UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

SARA PATRÍCIA DE OLIVEIRA SANTOS

CONSTRUÇÃO DE NOVAS MOLÉCULAS RECOMBINANTES HIPOALERGÊNICAS COM POTENCIAL APLICAÇÃO IMUNOTERAPÊUTICA UTILIZANDO ABORDAGENS DE BIOLOGIA SINTÉTICA

Salvador
2022

SARA PATRÍCIA DE OLIVEIRA SANTOS

CONSTRUÇÃO DE NOVAS MOLÉCULAS RECOMBINANTES HIPOALERGÊNICAS COM POTENCIAL APLICAÇÃO IMUNOTERAPÊUTICA UTILIZANDO ABORDAGENS DE BIOLOGIA SINTÉTICA

Tese apresentada ao Programa de Pós-Graduação em Biotecnologia, Instituto de Ciências da Saúde, Universidade Federal da Bahia como requisito para obtenção do título de Doutor em Biotecnologia.

Orientador: Prof. Dr. Luis Gustavo Carvalho Pacheco

Coorientadora: Profa. Dra. Neuza Maria Alcantara-Neves

Salvador
2022

Ficha catalográfica elaborada pelo Sistema Universitário de Bibliotecas (SIBI/UFBA),
com os dados fornecidos pelo(a) autor(a).

de Oliveira Santos, Sara Patrícia
CONSTRUÇÃO DE NOVAS MOLÉCULAS RECOMBINANTES HIPOALERGÊNICAS COM POTENCIAL APLICAÇÃO IMUNOTERAPÊUTICA UTILIZANDO ABORDAGENS DE BIOLOGIA SINTÉTICA / Sara Patrícia de Oliveira Santos. -- Salvador, 2022.
170 f. : il

Orientador: Luis Gustavo Carvalho Pacheco.
Coorientador: Neuza Maria Alcantara-Neves.
Tese (Doutorado - Doutorado em Biotecnologia) -- Universidade Federal da Bahia, Instituto de Ciências da Saúde, 2022.

1. Proteínas recombinantes. 2. Unidade Operacional de Expressão. 3. Biologia Sintética. 4. Imunoterapia alérgeno-específica. 5. Der p 21. I. Carvalho Pacheco, Luis Gustavo. II. Alcantara-Neves, Neuza Maria. III. Título.

SARA PATRÍCIA DE OLIVEIRA SANTOS

CONSTRUÇÃO DE NOVAS MOLÉCULAS RECOMBINANTES HIPOALERGÊNICAS COM POTENCIAL APLICAÇÃO IMUNOTERAPÊUTICA UTILIZANDO ABORDAGENS DE BIOLOGIA SINTÉTICA

Tese apresentada como requisito para obtenção do grau de Doutora em Biotecnologia pelo Instituto de Ciências da Saúde da Universidade Federal da Bahia.

Aprovada em 03 de novembro de 2022.

BANCA EXAMINADORA:

Luís Gustavo Carvalho Pacheco (orientador) _____
Doutorado em Bioquímica e Imunologia pela Universidade Federal de Minas Gerais, UFMG, Brasil.
Universidade Federal da Bahia.

Suzana Telles da Cunha Lima
Doutora em Biologia Vegetal pela Universidade Estadual de Campinas, UNICAMP, Brasil.
Universidade Federal da Bahia.

Silvana Beutinger Marchioro _____
Doutora em Faculty of Veterinary Sciences pelo Ghent University, UGENT, Bélgica.
Universidade Federal da Bahia.

Documento assinado digitalmente
gov.br
MARILIA OLIVEIRA FONSECA GOULART
Data: 09/11/2022 20:25:24-0300
Verifique em https://verificador.iti.br

Marilia Oliveira Fonseca Goulart _____
Doutora em Química pela Universidade Federal de Minas Gerais, UFMG, Brasil.
Universidade Federal de Alagoas.

Raquel Guimarães Benevides _____
Doutora em Bioquímica pela Universidade Federal do Ceará, UFC, Brasil.
Universidade Estadual de Feira de Santana.

AGRADECIMENTOS

À Deus, pelo dom da vida e por me oferecer tantas oportunidades e com elas obstáculos que me impulsionaram a crescer cada vez mais.

À minha família, que sempre torceu para que eu tivesse êxito tanto na minha vida profissional como pessoal. Agradeço imensamente por todo o apoio.

Ao meu orientador, Prof. Dr. Luis Gustavo Carvalho Pacheco, pelo incentivo e por todos os ensinamentos. Agradeço por acreditar no meu trabalho e contribuir para o meu desenvolvimento profissional. Serei sempre grata pela oportunidade.

À profa. Dra. Carina da Silva Pinheiro, que sempre esteve presente, enriquecendo o nosso trabalho com a sua experiência e sugestões. Além de ter se tornado uma pessoa muito especial durante toda essa jornada.

À profa. Dra. Neuza Maria Alcantara-Neves, por me receber em seu laboratório e aceitar ser minha coorientadora, realizando importantes contribuições na execução deste trabalho.

À equipe de trabalho do Laboratório de Alergia e Acarologia do Instituto de Ciências da Saúde (UFBA) pelos conhecimentos compartilhados e pela companhia diária, em especial à Fabián, Filipe e Leonardo, vocês sempre farão parte da minha vida.

Aos servidores do Instituto de Ciências da Saúde da UFBA e ao Programa de Pós-Graduação em Biotecnologia, por todo o suporte necessário.

Às agências de fomento, FAPESB, CNPq e CAPES, pelo financiamento deste trabalho.

E a todos aqueles que torceram por mim e que, de alguma forma, contribuíram para a conclusão deste trabalho.

"A gratidão é o único tesouro dos humildes." (William Shakespeare)

LISTA DE ABREVIATURAS

3D – Three-dimensional (inglês), tridimensional (português)

ΔΔG - Gibbs free energy (inglês), energia livre de Gibbs (português)

AA - Amino acid (inglês); Aminoácido (português)

AIT - Allergen-specific immunotherapy (inglês); Imunoterapia alérgeno-específica (português)

Asp(D) - Aspartic acid (inglês), ácido aspártico (português)

CDS - Coding sequence (inglês); Sequência codificadora (português)

DNA - Deoxyribonucleic Acid (inglês); Ácido desoxirribonucleico (português)

DO – Densidade óptica

DpE - *D. pteronyssinus extract* (inglês); extrato de *D. pteronyssinus* (português)

D. pteronyssinus - Dermatophagoides pteronyssinus

E. coli - Escherichia coli

ELISA – Enzyme-linked immunosorbent assay (inglês); Ensaio de imunoabsorção enzimática (português)

EOU - Expression Operational Unit

Gly(G) – Glycine (inglês); glicina (português)

Glu(E) - Glutamic acid (inglês); ácido glutâmico (português)

HDM - House dust mite (inglês); ácaro da poeira doméstica (português)

His – Histidine (inglês); histidina (português)

HRP - Streptavidin-horseradish peroxidase

IgA - Immunoglobulin A (inglês); Imunoglobulina A (português)

IgE - Immunoglobulin E (inglês); Imunoglobulina E (português)

IgG – Immunoglobulin G (inglês); Imunoglobulina G (português)

IgG4 – Imunoglobulina G4

IL-8 – Interleucina 8

IL-10 – Interleucina 10

IL-12 – Interleucina 12

INF-gamma(γ) – Interferon-gamma

IPTG – Isopropyl β-D-1-thiogalactopyranoside (inglês); Isopropil-β-D-tiogalactopirosídeo (português)
kDa – Kilodalton
LB - Luria-Bertani
Lys(K) – Lysine (inglês), lisina (português)
MDS - Molecular dynamics simulation (inglês); Simulação de dinâmica molecular (português)
mM – Milimolar
MHC - Major histocompatibility complex (inglês); Complexo principal de histocompatibilidade (português)
MW – Molecular weight (inglês); peso molecular (português)
NaP – Sodium phosphate (inglês); fosfato de sódio (português)
nm - Nanômetro
ns - Nanosegundo
OD – Optical density
OMS – Organização mundial da saúde
P1.D82P – Mutante 1
P2.K110G – Mutante 2
P3.E77G – Mutante 3
P4.E87S – Mutante 4
PBS – Phosphate buffered saline (inglês); salina tamponada com fosfato (português)
PDB – Protein data bank
pH - Potencial hidrogeniônico
Pro(P) – Proline (inglês); prolina (português)
ProAR - Bahia State Program for the Control of Asthma and Allergic Rhinitis (inglês); Programa de controle de asma e da rinite alérgica na Bahia (português)
RBS - Ribosome binding site (inglês); Sítio de ligação ao ribossomo (português)
RFU - Relative fluorescence units (inglês); Unidades de fluorescência relativa (português)
RMSD - Root mean squared deviation (inglês); Desvio Quadrático Médio (português)

RMSF - Root mean square fluctuations (inglês); Flutuação Quadrática Média (português)

RNA - Ribonucleic Acid (inglês); Ácido ribonucleico (português)

RT-qPCR – Reverse transcription-quantitative polymerase chain reaction (inglês); Reação em cadeia da polimerase quantitativa com transcrição reversa (português)

SDS-PAGE – Sodium dodecyl sulfate polyacrylamide gel electrophoresis

Ser(S) – Serine (inglês); serina (português)

*sf*GFP – *superfolder*GFP

TB - Terrific Broth

TGF-beta(β) – Fator de transformação do crescimento beta

TH1 – T helper 1 (inglês); T auxiliar 1 (português)

TH2 – T helper 2 (inglês); T auxiliar 2 (português)

UOE - Unidade Operacional de Expressão

WT - Wild-type (inglês); tipo-selvagem (português)

LISTA DE TABELAS

LISTA DE FIGURAS

SANTOS, Sara Patrícia de Oliveira. CONSTRUÇÃO DE NOVAS MOLÉCULAS RECOMBINANTES HIPOALERGÊNICAS COM POTENCIAL APLICAÇÃO IMUNOTERAPÊUTICA UTILIZANDO ABORDAGENS DE BIOLOGIA SINTÉTICA. 170 f. 2022. Tese de Doutorado - Instituto de Ciências da Saúde, Universidade Federal da Bahia, Salvador, 2022.

RESUMO

O desenvolvimento da biotecnologia moderna demanda quantidades satisfatórias de proteínas recombinantes solúveis e biologicamente ativas, tendo em vista a crescente utilização de proteínas heterólogas em laboratórios de pesquisa e aplicações terapêuticas. Portanto, a seleção de um sistema de expressão ideal, incluindo os elementos genéticos e o hospedeiro para expressão, é uma questão importante no desenvolvimento de proteínas recombinantes. Recentemente, foi desvendada a influência significativa que os elementos genéticos exercem na produção de uma proteína recombinante, sendo possível a otimização desses elementos graças aos avanços no campo da biologia sintética. Estudos de dados da literatura realizados pelo nosso grupo de pesquisa permitiram a montagem de uma nova Unidade Operacional de Expressão (UOE) composta por partes biológicas padronizadas e bem caracterizadas para melhorar a expressão de proteínas recombinantes em hospedeiros bacterianos. Também construímos uma nova variante do promotor T7 com atividade transcricional aumentada (1,7 vezes maior) em comparação com o promotor T7 tipo-selvagem. Essa nova UOE gerou uma produção melhorada da proteína repórter *superfolder*GFP (*sf*GFP) em *Escherichia coli* BL21(DE3) (unidades de fluorescência relativa/RFU = 70,62 ± 1,62 A U.), quando comparada a um vetor de expressão utilizado como controle (plasmídeo BBa_I746909; RFU = 59,68 ± 1,82 A U.). A funcionalidade dessa UOE também foi avaliada com a produção do alérgeno recombinante Der p 21 wt e suas variantes mutantes hipoalergênicas. O Der p 21 é considerado um alérgeno importante, com alta capacidade de ligação a anticorpos IgE e alta atividade alergênica. Isso define essa proteína como potencial candidata à imunoterapia alérgeno-específica (AIT) contra alergia desencadeada por *Dermatophagoides pteronyssinus.* Sendo assim, variantes hipoalergênicas do Der p 21 foram construídas para compor formulações de vacinas para AIT, seguindo uma abordagem de bioinformática estrutural e imunoinformática. Duas dessas variantes mutantes (K110G e E87S) apresentaram baixa capacidade de ligação à IgE em comparação com o alérgeno rDer p 21. Logo, essas novas proteínas hipoalergênicas são candidatas promissoras para compor formulações de imunoterapia alérgeno-específica de próxima geração em um futuro próximo. Além disso, o uso de vetores de expressão otimizados, como o montado neste estudo, oferece excelentes vantagens na produção de proteínas recombinantes de interesse biotecnológico, permitindo uma produção eficiente com melhor rendimento e qualidade.

Palavras-chave: Proteínas recombinantes, Unidade Operacional de Expressão, Biologia Sintética, Imunoterapia alérgeno-específica, Der p 21.

SANTOS, Sara Patrícia de Oliveira. CONSTRUCTION OF NEW HYPOALLERGENIC RECOMBINANT MOLECULES WITH IMMUNOTHERAPEUTIC APPLICATION POTENTIAL EMPLOYING SYNTHETIC BIOLOGY APPROACHES. 170 pp. 2022. Ph.D. Thesis - Institute of Health Sciences, Federal University of Bahia, Salvador, 2022.

ABSTRACT

The development of modern biotechnology demands satisfactory amounts of soluble and biologically active recombinant proteins, given the growing use of heterologous proteins in research laboratories and therapeutic applications. Therefore, selecting an ideal expression system, including the genetic elements and the host for expression, is a major issue in developing recombinant proteins. Recently, the significant influence that genetic elements have on the production of a recombinant protein has been unveiled, and it is feasible to optimize these elements thanks to advances in the synthetic biology field. Studies of literature data carried out by our research group allowed the assembly of a new Expression Operational Unit (EOU) comprising standardized and well-characterized biological parts to improve the expression of recombinant proteins in bacterial hosts. We also constructed a novel variant of the T7 promoter with increased transcriptional activity (1.7-fold higher) compared to the wild-type T7 promoter. This new EOU generated an improved production of the superfolderGFP reporter protein (sfGFP) in *Escherichia coli* BL21(DE3) (relative fluorescence units/RFU = 70.62 ± 1.62 A U.) when compared to an expression vector used as control (plasmid BBa_I746909; RFU = 59.68 ± 1.82 A U.). The functionality of this UOE was also evaluated with the production of the recombinant allergen Der p 21 wt and its hypoallergenic mutant variants. Der p 21 is considered an important allergen, with a high binding capacity for IgE antibodies and high allergenic activity. It defines this protein as a potential candidate for allergen-specific immunotherapy (AIT) against allergy triggered by *Dermatophagoides pteronyssinus*. Thus, hypoallergenic variants of the Der p 21 allergen were constructed to compose vaccine formulations for TIA, following a structural bioinformatics and immunoinformatics approach. Two of these mutants variants (K110G and E87S) displayed a low binding capacity to IgE compared to the allergen rDer p 21. Hence, these new hypoallergenic proteins are promising candidates for composing next-generation allergen-specific immunotherapy formulations in the near future. Moreover, the use of optimized expression vectors, such as the one assembled in this study, offers excellent advantages in the production of recombinant proteins of biotechnological interest, allowing for efficient production with improved yield and quality.

Keywords: Recombinant proteins, Expression Operational Unit, Synthetic Biology, Allergen-specific immunotherapy, Der p 21.

SUMÁRIO

1 INTRODUÇÃO

O uso de proteínas recombinantes para diversas aplicações da biotecnologia tem se tornado cada vez mais evidente, gerando benefícios tanto para o meio ambiente como para a saúde humana e animal (BHANDARI; LIM; GARDNER, 2021; CHEN, 2012; LOZANO TEROL et al., 2021; PUETZ; WURM, 2019; SANTOS et al., 2022a; TRIPATHI; SHRIVASTAVA, 2019; TUNGEKAR; CASTILLO-CORUJO; RUDDOCK, 2021). Porém, alguns obstáculos ainda são enfrentados na produção de proteínas recombinantes com a qualidade e rendimento de produção requeridos (BHANDARI; LIM; GARDNER, 2021; KOMAR, 2016; SCHLEGEL; GENEVAUX; DE GIER, 2017). Um dos principais desafios é a escolha de um sistema de expressão heterólogo ideal, o qual inclui os elementos genéticos e o hospedeiro para a expressão (GUSTAFSSON et al., 2012; KAUR; KUMAR; KAUR, 2018).

A seleção de um sistema inapropriado pode resultar em uma proteína com estrutura incorreta ou com baixos níveis de produção, e em casos mais críticos a produção não é alcançada (CHEN, 2012; KAUR; KUMAR; KAUR, 2018; PACHECO et al., 2012). Em muitas situações vários sistemas de expressão devem ser testados até que o melhor seja definido, o que pode tornar o processo de produção mais demorado e caro (PACHECO et al., 2012). Logo, o desenvolvimento de metodologias que permitem uma produção mais previsível e eficiente é essencial para diminuir os custos e garantir o sucesso da produção (MUTALIK et al., 2013a).

Vários hospedeiros de expressão de proteínas recombinantes já estão bem estabelecidos (DRAGOSITS; NICKLAS; TAGKOPOULOS, 2012; KOMAR, 2016; TRIPATHI; SHRIVASTAVA, 2019). Dentre eles, a bactéria *Escherichia coli* impulsionou a produção de proteínas recombinantes e continua sendo o hospedeiro mais utilizado (CHEN, 2012; LI et al., 2022; MARSCHALL; SAGMEISTER; HERWIG, 2016; PACKIAM et al., 2020; SILVA et al., 2021; TUNGEKAR; CASTILLO-CORUJO; RUDDOCK, 2021). Isso porque esse hospedeiro oferece um método simples e rápido para a expressão de proteínas recombinantes (LI et al., 2022; ROSANO; MORALES; CECCARELLI, 2019). Além disso, esse microrganismo possui uma genética bem caracterizada, diversas cepas mutantes disponíveis e várias das metodologias usadas para produção de

proteínas recombinantes estão otimizadas para *E. coli* (ROSANO; CECCARELLI, 2014). Isso reflete na estimativa de que aproximadamente 88% das estruturas proteicas depositadas no *Protein Data Bank* (PDB) são derivadas de proteínas produzidas neste microrganismo (SILVA et al., 2021). Além disso, a *E. coli* foi utilizada como hospedeira em cerca de 75% dos estudos acadêmicos envolvendo proteínas recombinantes (BILL, 2014), e também na produção de cerca de 30% das biomoléculas terapêuticas recombinantes pela indústria de biotecnologia (SILVA et al., 2021).

Com relação aos elementos genéticos envolvidos na produção de proteínas recombinantes, pouco entendimento existia sobre a otimização dos componentes genéticos dos sistemas de expressão, e todo o esforço empregado vinha sendo direcionado para poucos elementos regulatórios definidos (VOGL; HARTNER; GLIEDER, 2013; WELCH et al., 2009). Hoje, sabe-se que os elementos genéticos empregados para expressar uma proteína recombinante podem ter um efeito determinante no nível de expressão e na qualidade da produção (CORREA; OPPEZZO, 2011; ROSANO; CECCARELLI, 2014).

Neste contexto, avanços na área de biologia sintética proporcionam ferramentas para a análise sistemática das variáveis de sequências que afetam a expressão de proteínas (GUSTAFSSON et al., 2012; KAUR; KUMAR; KAUR, 2018). Com isso, o uso da biologia sintética permite que as sequências nucleotídicas das regiões codificadoras e dos seus elementos regulatórios, tais como promotores e sítios de início da tradução, sejam transformadas em parâmetros que podem ser otimizados para a expressão (DRAGOSITS; NICKLAS; TAGKOPOULOS, 2012; GUSTAFSSON et al., 2012; KAUR; KUMAR; KAUR, 2018; LOZANO TEROL et al., 2021; SANTOS et al., 2022a). Sendo assim é possível realizar modificações em vetores de expressão tradicionalmente usados (SADEGHIAN-RIZI et al., 2019; SHILLING et al., 2020) ou até mesmo construir vetores novos para permitir a produção ajustável das proteínas de interesse (FIRAT DUZENLI; OKAY, 2020; MIKIEWICZ et al., 2019; ZHANG et al., 2021). Seguindo essa linha de pesquisa, no capítulo 1 deste trabalho demonstramos a construção de uma nova Unidade Operacional de Expressão (UOE) para produção de proteína recombinante em *E. coli*, com base na combinação de vários elementos genéticos padronizados e bem caracterizados para transcrição e tradução melhorada de um gene

de interesse (SANTOS et al., 2022a).

Sobre as aplicações terapêuticas das proteínas recombinantes destaca-se a produção de vacinas para doenças respiratórias alérgicas, como a asma e a rinite, as quais ocorrem em todo o mundo e suas prevalências têm aumentado ao longo das últimas décadas, atingindo proporções epidêmicas (ANSOTEGUI et al., 2020; KHALAF et al., 2022; VALENTA; CAMPANA; NIEDERBERGER, 2017). De acordo com as estimativas da Organização Mundial da Saúde (OMS) em 2015, cerca de 400 milhões de pessoas foram afetadas pela rinite alérgica e sua prevalência continua aumentando (PAWANKAR, 2021). Estima-se que a asma afeta 262 milhões de pessoas no mundo, causando mortalidade significativa de 461.000 mortes anualmente (ABBAFATI et al., 2020; MEGHJI et al., 2021), e a previsão é de que em 2025 aproximadamente 400 milhões de pessoas sejam afetadas por essa doença (ANSOTEGUI et al., 2020).

Antigamente as doenças alérgicas eram consideradas doenças de países desenvolvidos, hoje ocorrem no mundo inteiro, inclusive na América Latina (MEGHJI et al., 2021). No Brasil, muitas capitais apresentam índices preocupantes para diversas doenças alérgicas (ALVES et al., 2020; DE ARAUJO CARDOSO et al., 2017; FERNANDES et al., 2017). Em Salvador vários indivíduos asmáticos foram registrados e têm sido atendidos pelo Programa de Controle de Asma e da Rinite Alérgica na Bahia (ProAR) (ALVES et al., 2020; CRUZ et al., 2010).

Diferentes estímulos ambientais são capazes de desencadear processos alérgicos, no entanto, um dos principais fatores de risco para esses processos é a sensibilização aos alérgenos dos ácaros da poeira doméstica (ACEVEDO; ZAKZUK; CARABALLO, 2019; CALDERÓN et al., 2015; CARVALHO et al., 2013; MILLER, 2019). *Dermatophagoides pteronyssinus* e *Dermatophagoides farinae* são as principais fontes de alérgenos de ácaros da poeira e estão disseminados em muitas regiões do globo, desencadeando sintomas alérgicos (ACEVEDO; ZAKZUK; CARABALLO, 2019; LIMÃO et al., 2020; SADE; ROITMAN; KIVITY, 2010). Dentre os alérgenos de *Dermatophagoides pteronyssinus*, Der p 21 foi caracterizado como um alérgeno importante, com alta capacidade de ligação a anticorpos IgE e alta atividade alergênica (PULSAWAT et al., 2014; WEGHOFER et al., 2008). Sendo considerado um potencial candidato não apenas para o diagnóstico de alergia mediada por IgE, mas também

para imunoterapia alérgeno-específica (AIT) contra alergia desencadeada por *D. pteronyssinus*.

A AIT é uma abordagem terapêutica alternativa ao uso de medicamentos para prevenir ou tratar as doenças alérgicas, utilizada há mais de 100 anos (AKDIS; AKDIS, 2015; CASALE; STOKES, 2014; DOROFEEVA et al., 2021). Esse tipo de terapia é baseada na administração de doses crescentes do alérgeno, ao qual o paciente possui reação de hipersensibilidade mediada por Imunoglobulina E (IgE), com o objetivo de estimular o seu sistema imunológico a desenvolver tolerância ao alérgeno e diminuir a sensibilidade específica contra o mesmo (AKDIS; AKDIS, 2015). Consequentemente, ocorre a melhora dos sintomas associados às alergias e promove uma proteção aos indivíduos contra o desenvolvimento de sintomas alérgicos (CASALE; STOKES, 2014; MAROGNA et al., 2010).

O método tradicional para AIT baseia-se na utilização de extratos brutos de alérgenos naturais, que consistem em misturas heterogêneas de alérgenos e não-alérgenos, provenientes de diferentes fontes (WALLNER; PICHLER; FERREIRA, 2013; YU; LIAO; TSAI, 2015). Uma das principais desvantagens do uso de extratos é o fato de que a composição molecular desses extratos pode variar consideravelmente entre as várias fontes e lotes, não permitindo o tratamento de acordo com o perfil de sensibilização individual do paciente. Além disso, eles precisam ser administrados durante longos períodos para alcançar resultados eficazes, e podem causar efeitos secundários graves em pacientes, incluindo o choque anafilático que pode levar à morte (FERREIRA; WOLF; WALLNER, 2014; SILVA et al., 2016, 2020; VALENTA; CAMPANA; NIEDERBERGER, 2017).

Para contornar essas desvantagens, os avanços na caracterização molecular de alérgenos levaram ao desenvolvimento de novas formas de AIT baseadas em proteínas recombinantes purificadas e derivados de peptídeos hipoalergênicos (AKINFENWA et al., 2021; SILVA et al., 2020). No entanto, a imunoterapia com alérgenos recombinantes purificados pode resolver os problemas relacionados ao uso de extratos de alérgenos mal definidos, mas não aborda a segunda grande desvantagem da imunoterapia seguindo essa metodologia, ou seja, também pode provocar efeitos secundários, variando de leve a local para sintomas mais graves (SILVA et al., 2016; VRTALA et al.,

2004). Sendo assim, para superar esses efeitos indesejáveis e para aumentar a eficácia, ferramentas modernas de engenharia de proteínas têm sido utilizadas em conjunto com novas ferramentas de bioinformática com o objetivo de desenvolver alérgenos hipoalergênicos (DA SILVA et al., 2020; THALHAMER et al., 2010; VALENTA; CAMPANA; NIEDERBERGER, 2017). Moléculas recombinantes hipoalergênicas são proteínas que possuem baixa ligação à IgE ao passo que seus epítopos de reconhecimento de células T e para indução de anticorpos IgG-tipo bloqueadores são preservados; logo, não provocam efeitos colaterais mas continuam capazes de gerar resposta imunológica de tolerância (SILVA et al., 2016). Desta forma, alérgenos recombinantes hipoalergênicos são considerados mais adequados e seguros, o que torna possível considerar o uso de vacinas de alérgenos para prevenção e tratamento da alergia (SILVA et al., 2016; VALENTA et al., 2011).

Cabe ressaltar que estudos com proteínas desenhadas *in silico* já demonstraram uma redução significativa na capacidade de ligação e reatividade cruzada de IgE, e manutenção das propriedades estimulantes de células T, comprovando o sucesso do uso de ferramentas de bioinformática para produção de moléculas hipoalergênicas (CHEN et al., 2014; LI et al., 2014; SIRCAR et al., 2016; THALHAMER et al., 2010). No capítulo 2 deste trabalho exploramos a metodologia de mutagênese *in silico* de proteínas alergênicas para produzir moléculas hipoalergênicas derivadas do alérgeno Der p 21 tipo-selvagem (wt) para o futuro desenvolvimento de vacinas de nova geração para doenças alérgicas causadas pelo ácaro *D. pteronyssinus.* Para isso foi utilizado um *pipeline* de análise de bioinformática estrutural, o qual permitiu o desenho racional de variantes hipoalergênicas, com redução significativa da reatividade à IgE (SANTOS et al., 2022b).

2 OBJETIVOS

2.1 OBJETIVO GERAL

O principal objetivo deste trabalho é construir novas moléculas recombinantes hipoalergênicas para imunoterapia alérgeno-específica para doenças respiratórias alérgicas, e otimizar a produção dessas moléculas em sistemas bacterianos.

2.2 OBJETIVOS ESPECÍFICOS

Objetivo 1: Construir e caracterizar uma nova Unidade Operacional de Expressão (UOE) direcionada por um promotor T7 otimizado, para melhorar a expressão de proteínas recombinantes de interesse biotecnológico em sistemas bacterianos.

- Construir e avaliar a interação de partes biológicas padronizadas para expressão de proteínas recombinantes em sistemas bacterianos;
- Construir uma nova variante otimizada do promotor T7;
- Inserir um novo módulo de clonagem na nova UOE, para que proteínas de interesse biotecnológico de difícil expressão sejam produzidas de maneira mais rápida e viável;
- Testar a funcionalidade da nova UOE em diferentes linhagens bacterianas, por meio da expressão da proteína repórter *superfolder*GFP (*sf*GFP).

Objetivo 2: Construir e avaliar novas moléculas recombinantes hipoalergênicas com potencial aplicação para imunoterapia alérgeno-específica (AIT) para doenças alérgicas causadas pelo ácaro *Dermatophagoides pteronyssinus.*

- Utilizar um *pipeline* de bioinformática estrutural para criar variantes hipoalergênicas do alérgeno Der p 21 do ácaro da poeira doméstica *Dermatophagoides pteronyssinus.*
- Utilizar o vetor de expressão pOPT 2.0, recém-criado pelo nosso grupo de pesquisa (SANTOS et al., 2022a), para produzir o alérgeno Der p 21

recombinante e suas variantes mutantes, utilizando *Escherichia coli* como hospedeiro;

- Purificar e avaliar a produção do alérgeno Der p 21 recombinante e das suas variantes mutantes;
- Caracterizar as novas moléculas hipoalergênicas produzidas.

3 CAPÍTULO 1: CONSTRUÇÃO E AVALIAÇÃO DE UMA NOVA UNIDADE OPERACIONAL DE EXPRESSÃO PARA PRODUÇÃO DE PROTEÍNAS RECOMBINANTES EM SISTEMAS BACTERIANOS

O surgimento da tecnologia do DNA recombinante possibilitou a expressão de proteínas a partir de células que não as produzem naturalmente; essas proteínas são conhecidas como proteínas recombinantes ou heterólogas (TRIPATHI; SHRIVASTAVA, 2019). A capacidade de se produzir proteínas recombinantes em diferentes sistemas heterólogos, com bom rendimento e qualidade, é um aspecto fundamental para o desenvolvimento da biotecnologia moderna, envolvendo aplicações médicas, acadêmicas e industriais (LOZANO TEROL et al., 2021; ROSANO; MORALES; CECCARELLI, 2019). Porém, durante o processo de produção podem ser encontrados obstáculos, sendo o principal desafio a escolha de um sistema de expressão heterólogo ideal, o qual inclui os elementos genéticos e o hospedeiro para a expressão (GUSTAFSSON et al., 2012; KAUR; KUMAR; KAUR, 2018).

As proteínas recombinantes são obtidas principalmente por fermentação microbiana, sendo a bactéria *Escherichia coli* o hospedeiro mais utilizado (LI et al., 2022; LOZANO TEROL et al., 2021; SILVA et al., 2021; TUNGEKAR; CASTILLO-CORUJO; RUDDOCK, 2021). Contudo, alguns problemas estão associados à produção de proteínas recombinantes em *E. coli*, como a toxicidade do produto, a formação de corpos de inclusão, a carga metabólica ou o sistema ineficiente de translocação/transporte de proteínas expressas (LOZANO TEROL et al., 2021; MARSCHALL; SAGMEISTER; HERWIG, 2016). Dessa forma, a otimização da transcrição e tradução de genes heterólogos é essencial para evitar esses inconvenientes e desenvolver uma produção eficiente (LOZANO TEROL et al., 2021).

Nos últimos anos, esforços estão sendo realizados para melhorar a expressão de proteínas recombinantes em sistemas bacterianos, e dentre eles destaca-se a busca de variáveis que interferem na sua produção (GUSTAFSSON et al., 2012; KAUR; KUMAR; KAUR, 2018; LOZANO TEROL et al., 2021). Neste contexto, os elementos genéticos possuem uma grande influência na produção de proteínas recombinantes, sendo possível otimizar esses elementos graças aos avanços na área da biologia sintética

(KAUR; KUMAR; KAUR, 2018). Além de permitir a edição dos elementos genéticos envolvidos na síntese de proteínas, as ferramentas de biologia sintética também possibilitam a criação de novos dispositivos (SANTOS et al., 2022a; YOUNG; ALPER, 2010; ZHANG et al., 2021). Com o uso da biologia sintética é possível projetar e construir novas funções biológicas através da descoberta, caracterização e reaproveitamento de partes moleculares (KHALIL; COLLINS, 2010; TAN et al., 2021). Contudo, essa área permite a síntese artificial do material genético, o que torna a construção e otimização de sistemas biológicos mais fácil e confiável (KAHL; ENDY, 2013).

Vale destacar que para a obtenção de um aumento mais considerável do nível de expressão, não é ideal que os elementos genéticos sejam tratados isoladamente. Deve-se considerar a unidade operacional de expressão de uma proteína recombinante, a qual abrange os elementos genéticos envolvidos na expressão de um gene e as junções e limites que relacionam as partes (MUTALIK et al., 2013a). Neste capítulo será apresentado um estudo onde foram utilizadas abordagens de biologia sintética para construir uma nova Unidade Operacional de Expressão (UOE) apresentando uma combinação de elementos genéticos altamente padronizados e bem caracterizados, com a finalidade de melhorar a produção de proteínas recombinantes em hospedeiro bacteriano (SANTOS et al., 2022a). Adicionalmente, foi construída uma nova variante otimizada do promotor T7, contendo combinações de alterações na sequência do promotor T7 tipo-selvagem. Para testar a funcionalidade dessa nova UOE ela foi clonada em diferentes plasmídeos para gerar os vetores de expressão pOPT 1.0 e pOPT 2.0, cujas eficiências em expressar uma proteína de interesse foram então avaliadas usando a proteína fluorescente *superfolder*GFP (*sf*GFP) como repórter.

3.1 ARTIGO 1: ENGINEERING AN OPTIMIZED EXPRESSION OPERATING UNIT FOR IMPROVED RECOMBINANT PROTEIN PRODUCTION IN *ESCHERICHIA COLI*

Artigo publicado: *Protein Expr Purif*. (2022) 199:106150. doi: 10.1016/j.pep.2022.106150.

Engineering an Optimized Expression Operating Unit for Improved Recombinant Protein Production in *Escherichia coli*

Sara P. O. Santos[1], Luis Fabian S. Garcés[1,#], Filipe S. R. Silva[1,#], Leonardo F. Santiago[1], Carina S. Pinheiro[1], Neuza M. Alcantara-Neves[1], Luis G. C. Pacheco[1,*]

[1] *Institute of Health Sciences, Federal University of Bahia, Salvador-BA, Brazil.*

***Correspondence:** L. Pacheco. Department of Biotechnology, Institute of Health Sciences, Federal University of Bahia, Salvador-BA, Brazil. E-mail: luis.pacheco@ufba.br.
Present addresses: L.F.S.G.: Technical University of Ambato, Ecuador. F.S.R.S.: University of São Paulo, São Carlos-SP, Brazil.

ABSTRACT

Common strategies to improve recombinant protein production in *Escherichia coli* often involve the test and optimization of several different variables, when using traditional expression vectors that are commercially available. Now, modern synthetic biology-based strategies allow for extensive modifications of these traditional vectors, or even construction of entirely new modular vectors, so as to permit tunable production of the recombinant proteins of interest. Herein, we describe the engineering of a new expression operating unit (EOU; 938 bp) for producing recombinant proteins in *E. coli*, through the combinatorial assembly of standardized and well-characterized genetic elements required for transcription and translation (promoter, operator site, RBS, junction RBS-CDS, cloning module, transcriptional terminator). We also constructed a novel T7 promoter variant with increased transcriptional activity (1.7-fold higher), when compared to the canonical wild type T7 promoter sequence. This new EOU yielded an improved production of the reporter protein superfolder GFP (sfGFP) in *E. coli* BL21(DE3) (relative fluorescence units/ RFU = 70.62 ± 1.62 A.U.) when compared to a high-producing control expression vector (plasmid BBa_I746909; RFU = 59.68 ± 1.82 A.U.). The yields of purified soluble recombinant sfGFP were also higher when using the new EOU (188 $mg.L^{-1}$ culture vs. 108 $mg.L^{-1}$ in the control) and it performed similarly well when inserted into different plasmid backbones (pOPT1.0/Amp^R and pOPT2.0/Cm^R).

Keywords: *Escherichia coli,* expression operating unit, recombinant protein, synthetic biology.

1. Introduction

Recombinant protein production in bacteria is essential for basic research in the life sciences and for applications in the therapeutic and industrial fields (BHANDARI; LIM; GARDNER, 2021; FIRAT DUZENLI; OKAY, 2020; LOZANO TEROL et al., 2021; PUETZ; WURM, 2019; TRIPATHI; SHRIVASTAVA, 2019; TUNGEKAR; CASTILLO-CORUJO; RUDDOCK, 2021). *Escherichia coli* is the microbial workhorse for the production of heterologous proteins and remains as the most widely used expression host (CHEN, 2012; FIRAT DUZENLI; OKAY, 2020; KI; PACK, 2020; LI et al., 2022; MARSCHALL; SAGMEISTER; HERWIG, 2016; PACKIAM et al., 2020; ROSANO; MORALES; CECCARELLI, 2019; SILVA et al., 2021; TRIPATHI; SHRIVASTAVA, 2019; TUNGEKAR; CASTILLO-CORUJO; RUDDOCK, 2021). It is estimated that *ca*. 88% of protein structures deposited in the Protein Data Bank are derived from proteins produced in this host organism (SILVA et al., 2021). In addition, *E. coli* is utilized as a host in about 75% of the academic studies involving recombinant proteins (BILL, 2014), and also in the production of nearly 30% of the recombinant therapeutic biomolecules by the biotechnology industry (SILVA et al., 2021).

To circumvent difficulties commonly associated with recombinant protein production in bacteria, such as low protein yield and aggregation, different strategies are often employed that involve optimization of culturing conditions, and protocols to improve protein solubility and purification (BHANDARI; GARDNER, 2021; DANA, 2018; KONDO; YUMURA, 2020; LOZANO TEROL et al., 2021; SCHLEGEL; GENEVAUX; DE GIER, 2017). Additionally, several variables are traditionally optimized to improve heterologous protein production in *E. coli*, including the test of multiple genetically engineered host strains and the selection of the most appropriate vector among a range of commercially available expression plasmids, such as the families pET, pBAD and pTrc (AHMAD et al., 2018; KAUR; KUMAR; KAUR, 2018; ZHANG et al., 2021).

However, modern strategies to produce recombinant proteins in *E. coli* have involved either modifications in these traditionally used expression vectors (SADEGHIAN-RIZI et al., 2019; SHILLING et al., 2020) or the construction of entirely new modular vectors so as to permit tunable production of the proteins of interest

(FIRAT DUZENLI; OKAY, 2020; MIKIEWICZ et al., 2019; ZHANG et al., 2021). These strategies rely on recent advances in the synthetic biology field, that allow for better design, selection, synthesis, combination, and systematic analysis of new genetic elements, then improving expression vectors and recombinant protein production (GUSTAFSSON et al., 2012). This synthetic biology-guided construction of improved genetic elements has already been evaluated in the optimization of promoter elements (CHIZZOLINI et al., 2014; FIRAT DUZENLI; OKAY, 2020; PAUL et al., 2013; ZHOU et al., 2017), regulatory operator sites (DARBY; HINE, 2005), ribosome binding sites (RBS) (NA; LEE, 2010; NA; LEE; LEE, 2010; SALIS, 2011; SALIS; MIRSKY; VOIGT, 2009), coding sequences (CDS) (GUSTAFSSON et al., 2012; GUSTAFSSON; GOVINDARAJAN; MINSHULL, 2004; MENZELLA, 2011; PUIGBÒ et al., 2007; SCHLESINGER et al., 2017), and of transcription/ translation initiation regions, such as the junctions between untranslated 5' region (5'-UTR) / RBS / CDS (BERG et al., 2009; MIRZADEH et al., 2015; SEO et al., 2013). All these studies described extensively characterized and standardized biological parts that can now be easily reused to construct new improved genetic tools (KOSURI et al., 2013; MUTALIK et al., 2013a, 2013b).

In this study, we constructed a novel Expression Operating Unit (EOU) for recombinant protein expression in *E. coli* based on the combination of several well-characterized genetic elements for improved transcription and translation of a gene of interest (Fig. 1A). Additionally, a new optimized T7 promoter variant was constructed that combined sequence alterations previously shown to improve transcription by separate studies (Fig. 1B) (CHIZZOLINI et al., 2014; PAUL et al., 2013). This new EOU was cloned in different plasmid backbones to generate the expression vectors pOPT1.0 and pOPT2.0, whose efficiencies to express a protein of interest were then evaluated using the superfolder green fluorescent protein (sfGFP) as a reporter.

2. Materials and methods

2.1. Selection of standardized genetic elements and design of the Expression Operating Unit (EOU)

The construction of the new Expression Operating Unit (EOU) was based on tried-and-tested standardized genetic parts retrieved from the Registry of Standard Biological Parts (RSBP, http://parts.igem.org/Main_Page) or from selected studies through literature search (Supplementary Table S1). The constructed EOU (Fig. 1A) is driven by a modified T7 promoter variant, which carries a combination of mutations previously shown to render increased transcription efficiency (CHIZZOLINI et al., 2014; PAUL et al., 2013). Basically, four nucleotides were mutated in the internal sequence of the wild-type T7 promoter, the extension CCGGT was inserted in the 5’ end, and the 3’ terminus remained unchanged (Fig. 1B).

Besides the modified T7 promoter variant, the EOU was composed of: two consecutive lac operators (*lacOs* and *lacO3*), such as in reference (DARBY; HINE, 2005); a strong ribosome binding site (RBS) (BBa_B0034); a well-defined RBS-CDS junction sequence, as described in (MIRZADEH et al., 2015); a cloning module containing the coding sequence for the superfolder GFP reporter protein (*sfgfp* – 720 bp) flanked by restriction sites for the type IIS enzyme *SapI*, for streamlined CDS replacement; a region coding for an in-frame 6x histidine tag; and a T7 terminator site (BBa_B1006). To standardize the cloning steps, the prefix and suffix sequences for the RFC23 cloning standard were added to the 5' and 3' ends of the EOU, respectively (RØKKE et al., 2014) (Fig. 1C; Supplementary Fig. S1).

The EOU (938 bp) was firstly inserted in pUC57, to yield the vector pOPT 1.0 (3.65 Kbp; pUC_ori/Amp^R) (Supplementary Fig. S2). The EOU was also sub-cloned in the pSB1C3 backbone, through digestion with *EcoRI* and *PstI* to generate the vector pOPT 2.0 (Supplementary Fig. S2; Supplementary Table S2).

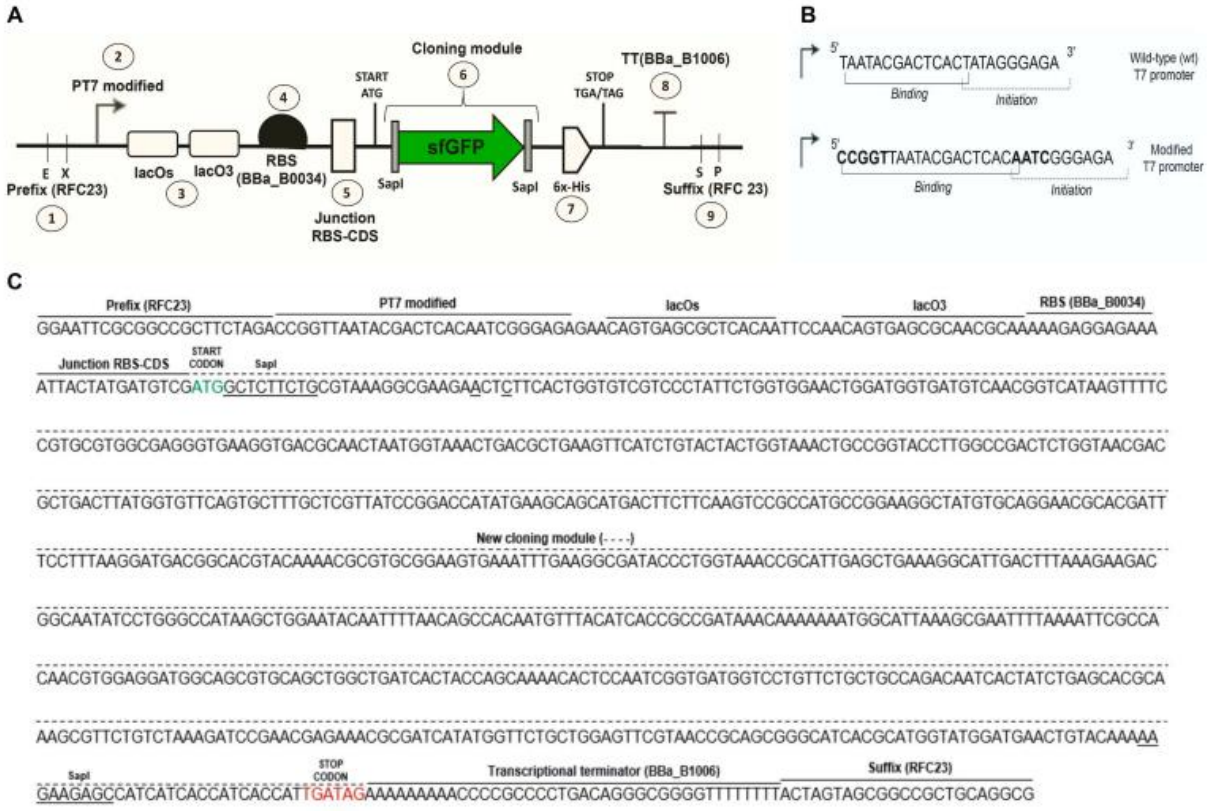


Figura 1 (tese). Fig. 1. Schematic of the engineered standardized Expression Operating Unity (EOU). (A) Standard genetic elements that compose the EOU: 1- standard prefix RFC 23; 2- modified T7 promoter variant; 3- operator sites (lacOs and lacO3); 4- strong RBS (BBa_B0034); 5- RBS-CDS junction; 6- Type IIS-based cloning module; 7- 6xHis C-terminal tag; 8- transcriptional terminator (BBa_B1006); 9- standard suffix RFC 23. (B) Sequence comparison between the modified T7 promoter variant and the wild-type T7 promoter. Alterations are shown in bold. (C) Sequence map of the EOU.

2.2. Protein expression standardization using the new EOU

The protein sfGFP was used as a reporter to standardize the recombinant protein expression from the pOPT 1.0 vector transformed in different *E. coli* strains (Supplementary Table S3). Measurement of sfGFP fluorescence in the culture was initially used as an indicative of total protein production. The plasmid BBa_I746909, obtained from the iGEM Registry of Standard Biological Parts (http://parts.igem.org/Part:BBa_I746909), was used as a positive control. This plasmid carries the coding sequence for the sfGFP protein driven by the canonical wild-type T7 promoter, a strong RBS sequence (BBa_B0034), and two transcriptional terminators (B0010 and B0012) (Supplementary Table S2; Supplementary Fig. S3). Fluorescence

from *E. coli* cultures was monitored by Fluoroskan Ascent Microplate Fluorometer (Thermo Fisher) (Ex. = 485nm/Em. = 538nm).

2.3. Recombinant protein purification and analysis

To effectively quantify recombinant protein production, his-tagged sfGFP was purified from induced *E. coli* cultures. Firstly, bacterial extracts were analyzed by 12% SDS-PAGE and Western blot (mouse anti-His IgG, 1:4000). Protein solubility was evaluated by a standard protocol, as previously described (SANTOS et al., 2022b). The MagneHis Protein Purification System (Promega) was used for affinity purification of sfGFP. Total yields of recombinant sfGFP were quantified with the Qubit Protein Assay Kit (Thermo Fisher).

2.4. Transcriptional activity from the modified T7 promoter variant

Reverse-transcription quantitative real-time PCR was used to evaluate the transcriptional activity arising from the modified T7 promoter variant, in comparison to the wild-type T7 promoter. For this, we first selected from the iGEM Parts Registry a positive control plasmid (BBa_K567018) that renders a well-characterized and robust expression of a gene of interest (*egfp*) under the control of the canonical wild-type T7 promoter. We then constructed the pT7var plasmid (Supplementary Table S2) carrying an *egfp* expression cassette that has the same genetic structure from BBa_K567018 but changing the wt T7 promoter by the new variant. Both plasmids were transformed in *E. coli* BL21(DE3) and *egfp* transcription was stimulated with 1mM IPTG, at OD600nm = 0.3. Total RNA was isolated and then used in RT-qPCR experiments targeting *egfp* (Supplementary Methods).

2.5. Statistical analyses

GraphPad Prism software was used to perform statistical analyses. Normality of tested by the Kolmogorov-Smirnov and Shapiro-Wilk tests. ANOVA was used for variance analysis, with Bonferroni post-hoc test (95% significance level).

3. Results and discussion

Research efforts for enhancing the expression levels of recombinant proteins in bacterial hosts have been crucial for biomedical and biotechnological studies (LOZANO TEROL et al., 2021; TUNGEKAR; CASTILLO-CORUJO; RUDDOCK, 2021). Such efforts now involve the application of synthetic biology strategies to the development of standardized and well-characterized genetic elements, so-called biological parts, aiming at improved transcriptional and translational rates of genes of interest in bactéria (FIRAT DUZENLI; OKAY, 2020; KAHL; ENDY, 2013; MUTALIK et al., 2013a, 2013b; ZHANG et al., 2021). Herein, we assembled an optimized Expression Operating Unit (EOU) for recombinant protein production in *Escherichia coli* through reuse of some of these well-characterized biological parts that are currently available (Fig. 1A and 1C). This strategy of rationally designing new EOUs by leveraging extensively characterized genetic parts has already been shown to be an interesting alternative to avoid unnecessary trial and error work when trying to improve expression of a given protein of interest in bacteria (KWOK, 2010; MUTALIK et al., 2013a, 2013b).

Additionally, we created a new T7 promoter variant with enhanced transcriptional activity (Fig. 1B) through combination of previously tried and tested sequence changes. (CHIZZOLINI et al., 2014; PAUL et al., 2013). Particularly, Paul and collaborators (PAUL et al., 2013) created and screened a randomized T7 promoter library and selected mutants with high *in vitro* transcription activities. This study reported a T7 variant possessing substitutions towards the 3’-end of the sequence, which was able to improve the recombinant protein yield after translation. The study by Chizzolini and collaborators (CHIZZOLINI et al., 2014), in turn, achieved enhanced transcriptional activity from the wt T7 promoter through a 5 bp sequence extension (CCGGT) in the 5’-end. In our study, the combination of both sequence changes to form a modified T7 promoter variant (Fig. 1B) resulted in an additional 1.7-fold increase in transcriptional activity, when comparing by RT-qPCR the expression of the *egfp* gene from the plasmids pT7var and BBa_K567018 (Supplementary Table S4); this improvement was also translated into higher yields of purified sfGFP. Interestingly, in a recent study by Shilling and collaborators (SHILLING et al., 2020), it was reported that the truncated T7 promoter

sequence present in the family of vectors of the pET expression system renders a three-fold reduction in the expression of the sfGFP protein, when compared to a consensus wild-type T7 promoter, which is four bases longer. The study suggests that the truncation in the T7 sequence found in pET arises from a design flaw that reduces protein production (SHILLING et al., 2020). In our study, that was developed concomitantly, we extended this knowledge by demonstrating that the addition of tried and tested sequence extensions to the consensus T7 promoter sequence can further improve the expression of a protein of interest in *E. coli*.

Different studies (CORREA; OPPEZZO, 2011; TAN; HSIANG; NG, 2021) described the importance of using a T7 expression system with genetic elements that attenuate the toxicity generated by the overexpression that can be caused by this system. In this way, we created an expression vector with elements so that the expression was highly controlled. We used two consecutive lac operators (*lacOs* and *lacO3*) (DARBY; HINE, 2005), which together have a higher affinity for the *LacI* repressor protein, and a transcriptional terminator (BBa_B1006) with high transcription termination rate (~90%).

Gustafsson and collaborators (GUSTAFSSON et al., 2012) took a more general approach and tested different experimental variables to improve the yield of recombinant proteins (bacterial strain, temperature, aeration, nutrients, pH). They also considered the genetic elements that impact the recombinant protein production, including the coding sequence (CDS) itself and the genetic elements that compose the expression vector. Following a similar reasoning, we tested different variables to find the ideal culture conditions to maximize the yield of the sfGFP recombinant protein produced from the new EOU (Supplementary Table S3). The best sfGFP (~28 kDa) expression in our study was achieved in the *E. coli* BL21(DE3) strain grown in Luria-Bertani (LB), at 37°C for 20 hours, with agitation at 200 rpm and induction by 1mM of IPTG at OD_{600nm} = 0.3-0.6 (Fig. 2A and 2B). Most of the recombinant protein was produced in the soluble (Sol.) fraction after 20 h of induction, with only trace amounts detected at the insoluble (Insol.) fraction (Fig. 2A), even by Western blot analysis (Fig. 2B). The fluorescence emitted by the presence of soluble sfGFP in the bacterial cell extracts also corroborated the results observed by SDS-PAGE analysis (Fig. 2A, lower panel).

We conducted the experiments under the same conditions with the pOPT 1.0 plasmid, harboring the new EOU, and the positive control plasmid BBa_I746909. Then, we registered the fluorescence emission by the cell cultures over time, following induction (Fig. 2C; Supplementary Fig. S4). After twenty hours, there was a significantly higher sfGFP production from cultures carrying the new EOU (mean sfGFP fluorescence = 70.62 ± 1.62 A.U.), when compared to the positive control plasmid (mean sfGFP fluorescence = 59.68 ± 1.82 A.U.) (Fig. 2C; Supplementary Fig. S5). This result was also reproducible when the EOU was sub-cloned in a different plasmid backbone, to create pOPT2.0 (Fig. 2D). The construction of pOPT2.0 was also important to permit the subsequent cloning of different target genes in the EOU, as we identified an additional *SapI* restriction site outside the EOU in the pOPT1.0 backbone when attempting to digest that plasmid.

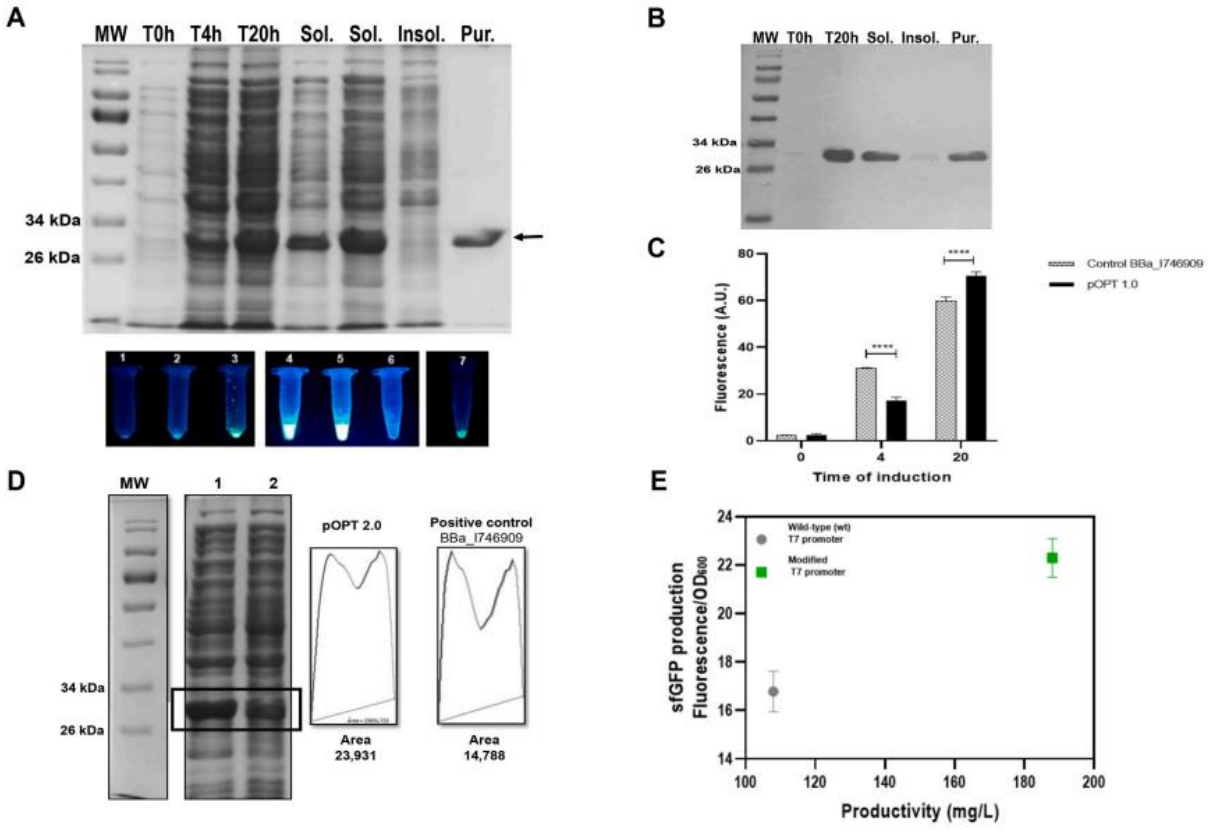


Figura 2 (tese). Fig. 2. Functioning of the new EOU evaluated through expression of recombinant sfGFP. (A) Upper panel, 12% SDS-PAGE showing expression of the sfGFP protein (ca. 28 kDa) in *E.coli*

BL21(DE3) transformed with the pOPT 1.0 vector. Time points (0h, 4h, 20h) represent period of induction by IPTG. Sol. = soluble fraction; Insol.= insoluble fraction. Pur.= purified proteins. Lower panel, sfGFP fluorescence emitted by the *E. coli* cell extracts at each matched sample in the SDS-PAGE: 1- T0h; 2- T4h; 3- T20h; 4- Sol.1; 5- Sol.2; 6- Insol.; 7- Pur. (B) Western blot analysis of sfGFP expression in the BL21(DE3) strain transformed with the pOPT1.0 vector. (C) sfGFP fluorescence from BL21(DE3) cultures transformed with either the pOPT 1.0 plasmid or the positive control plasmid BBa_I746909, after IPTG induction. Data is presented as mean ± S.D. (n = 5). **** Significantly different ($p < 0.0001$). (D) SDS-PAGE and densitometric analysis by ImageJ of the soluble recombinant sfGFP produced by *E. coli* transformed with the pOPT 2.0 plasmid or the positive control BBa_I746909. (E) Yields of soluble recombinant sfGFP in *E. coli* BL21(DE3) transformed with the pOPT 1.0 (modified P_{T7}) or the positive control BBa_I746909 (wild-type P_{T7}). sfGFP was purified to completion by affinity chromatography and the final yields were plotted against the sfGFP fluorescence of the culture normalized by the optical density at 600 nm.

Recombinant sfGFP was purified to completion by affinity chromatography (Fig. 2A and 2B), rendering final soluble protein yields of 188 $mg.L^{-1}$ culture from pOPT 1.0 and 108 $mg.L^{-1}$ culture from the positive control BBa_I746909 (Fig. 2E).

Producing high levels of recombinant proteins in bacteria in their soluble form can be an infrequent outcome, assuming the contradictory dynamics between these two variables in the recombinant protein expression context. (DRAGOSITS; NICKLAS; TAGKOPOULOS, 2012; SILVA et al., 2021). Obtaining biologically active and well-folded proteins is a bottleneck in many recombinant protein processes, having a huge demand in many industries (FENG et al., 2014; HON et al., 2021; MEE; BANKI; WOOD, 2008). Then, developing expression systems that prioritize these qualities should be the starting point of any project aiming to produce proteins of biotechnological interest.

In another recent study by our group, we already explored the effectiveness of our newly engineered EOU to express a protein of biotechnological interest, the house dust mite allergen Der p 21 (SANTOS et al., 2022b). In previous attempts to express this protein from commercially available vectors, such as pD444, we couldn't achieve a successful expression in *E. coli,* even when different expression conditions were tested (data not shown). Noteworthy, after cloning in the pOPT 2.0 plasmid, carrying the new EOU, the Der p 21 allergen and its mutant derivatives could be produced in *E. coli* at final soluble protein yields ranging from 55.6 to 60 mg.L-1 of culture (SANTOS et al., 2022b). The difficulties associated with production of soluble recombinant proteins from eukaryotic origin using traditional commercial vectors was also demonstrated by Pacheco and collaborators (PACHECO et al., 2012) when testing various target genes in

distinct *E. coli* strains, using a pET expression vector (pET28a-LIC) (PACHECO et al., 2012). Therefore, engineering new optimized expression vectors using well-characterized genetic elements may be a good alternative to tackle the problem of obtaining recombinant proteins in *E. coli* at desirable quantity and quality.

4. Conclusion

In this study we engineered a new expression operating unit (EOU) for producing recombinant proteins in *E. coli*, through the combination of standardized and well-characterized genetic elements. We also constructed a novel T7 promoter variant with increased transcriptional activity, when compared to the canonical wild type T7 promoter sequence. The new EOU yielded an improved production of a sfGFP in *E. coli*, mostly in the soluble form, and performed similarly well in different plasmid backbones. The ability to produce a hard-to-express protein from eukaryotic origin was already demonstrated by production of the Der p 21 allergen from house dust mite (SANTOS et al., 2022b). This study demonstrates the utility of reusing deeply characterized genetic elements in new combinations to develop optimized expression devices capable of improving recombinant protein production in bacteria.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgements

S.P.O.S. was recipient of a PhD scholarship from CAPES foundation, Brazil. N.M.A.N, and L.G.C.P. are recipients of research fellowships from CNPq-Brazil. This study was partially funded by FAPESB, through research grant #PNE0007/2014.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pep.2022.106150.

Supplementary Material

Engineering an Optimized Expression Operating Unit for Improved Recombinant Protein Production in *Escherichia coli*

Sara P. O. Santos[1], Luis Fabian S. Garcés[1], Filipe S. R. Silva[1], Leonardo F. Santiago[1], Carina S. Pinheiro[1], Neuza M. Alcantara-Neves[1], Luis G. C. Pacheco[1,*]

[1] *Institute of Health Sciences, Federal University of Bahia, Salvador-BA, Brazil.*
***Corresponding author:** Prof. Luis G. C. Pacheco. Department of Biotechnology, Institute of Health Sciences, Federal University of Bahia, Salvador-BA, 40.110-100, Brazil. E-mail: luis.pacheco@ufba.br

List of Contents

Supplementary Methods

Comparative analysis of the transcriptional activities of the wild-type T7 promoter and the new T7 promoter variant, by RT-qPCR

E. coli BL21(DE3) were separately transformed with the plasmid pT7var, containing the new T7 promoter variant, and the positive control plasmid BBa_K567018, containing the canonical T7 promoter consensus sequence (see Supplementary Table S2). The expression of the *egfp* gene, coding for enhanced GFP, was then induced by the addition of 1mM IPTG to the *E. coli* cultures, at an OD600nm = 0.3. Total RNA was isolated with the SV Total RNA Isolation System (Promega), reverse transcribed with the GoScript™ Reverse Transcription System (Promega), and then used in RT-qPCR experiments targeting *egfp* with the GoTaq® qPCR Master Mix (Promega).

Tabela 1 (tese). Supplementary Table S1. Standard genetic elements that compose the new expression operating unit (EOU)

Genetic elements	Sequence	Characteristics	Source
Modified T7 promoter	CCGGTTAATACGACTCACAATCGGGAGA	T7 promoter built from studies by Paul et al. (2013) and Chizzolini et al. (2014), in order to increase the transcriptional level	This study
lac operators - lacOs e lacO3	CAGTGAGCGCTCACAA /CAGTGAGCGCAACGCAA	Operator sites originated from the lacO operon, together have a greater affinity for the LacI repressor protein	Darby e Hine (2005)
Strong RBS (BBa_B0034)	AAAGAGGAGAAA	Ribosome binding site with strong affinity for *E. coli* 16S ribosomal RNA.	Mahajan et al. (2003) - Grupo: iGEM Warsaw 2010 (BioBricks Foundation)
Junction RBS-CDS	ATTACTATGATGTCG	Contributes to higher levels of expression.	Mirzadeh et al. (2015)
SapI-based cloning module[(a)]	**ATG**GCTCTTCT/G …**GOI**…/AAGAAGAGC**TGATAG**	New cloning module containing the sfGFP coding sequence flanked by sites of the SapI type IIS enzyme, which allows the facilitated replacement of the CDS.	This study
6x-His tail	CATCATCACCATCACCAT	6x-His tail to allow protein purification by affinity chromatography	
Transcriptional terminator (BBa_B6100)	AAAAAAAAACCCCGCCCCTGACAGGGCG GGGTTTTTTTT	High termination rate: ~ 90%	Huang (2006) – BioBricks Foundation

[(a)] GOI: gene of interest. The start codon and stop codon are already present in the plasmid, upstream and downstream from the cloning sites (marked in bold). An additional 'G' is inserted at the end of the *SapI* site for in-frame cloning.

Tabela 2 (tese). Supplementary Table S2. List of bacterial strains and plasmids used in this study.

	Characteristics	Source
Strains		
E. coli XL1Blue	*endA1 gyrA96(nalR) thi-1 recA1 relA1 lac glnV44 F' [:: Tn10 proAB+ lacIq Δ(lacZ)M15] hsdR17(rK- mK+)*	Phoneutria
E. coli BL21(DE3)	*F– ompT gal dcm lon hsdS$_B$(r_B–m_B–) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]$_{K-12}$(λ^S)*	NEB
E. coli BL21(DE3)pLysS	*F– ompT gal dcm lon hsdS$_B$(r_B–m_B–) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]$_{K-12}$(λ^S) pLysS[T7p20 ori$_{p15A}$](Cm^R)*	NEB
E. coli C43(DE3)	*F– ompT gal dcm hsdS$_B$(r_B- m_B-)(DE3)*	Invitrogen
Plasmids		
pOPT 1.0	pUC-derived pMB1 ori/ Amp^R; Expression vector containing an optimized T7 promoter, strong RBS, fragment encoding sfGFP, 6xHis-tag, T7 term. *SapI*-mediated scarless cloning.	This study
pOPT 2.0	pUC-derived pMB1 ori/ Cm^R; Expression vector containing an optimized T7 promoter, strong RBS, fragment encoding sfGFP, 6xHis-tag, T7 term. *SapI*-mediated scarless cloning.	This study
pT7var	pUC-derived pMB1 ori / Amp^R; Expression vector containing an optimized T7 promoter and the enhanced EGFP protein coding sequence	This study
pSB1C3	pUC-derived pMB1 ori/ Cm^R; Cloning vector.	iGEM Parts Registry
BBa_I746909	pUC-derived pMB1 ori/ Cm^R; Expression vector containing a wild type T7 promoter and the sfGFP protein coding sequence	iGEM Parts Registry
BBa_K567018	pUC-derived pMB1 ori/ Cm^R; Expression vector containing a wild type T7 promoter and the EGFP protein coding sequence	iGEM Parts Registry

Tabela 3 (tese). Supplementary Table S3. Variables tested for sfGFP protein expression from the new EOU

Variable	Conditions
E. coli strains	BL21(DE3), BL21(DE3)pLysS, and C43(DE3)
Media composition	LB (Luria-Bertani) or TB (Terrific Broth)
DO_{600nm} of induction	0.1 to 1.3
IPTG concentration	1mM, 1.5mM, 2.0mM
Growth temperature	16°C, 25°C and 37°C
Induction period	4 hours or 20 hours

Tabela 4 (tese). Supplementary Table S4. RT-qPCR results comparing the transcriptional activity from the new T7 variant (plasmid pT7var) and the wild-type T7 promoter (plasmid BBa_K567018)

	Target gene (*egfp*) Mean Ct ± S.D.	**Reference gene (*dnaG*) Mean Ct ± S.D.**	**ΔΔCt**	**Fold Change (relative to control)**
BL21(DE3) (pT7var)	11.39±0.04	21.88±0.68	**-0.78**	**1.7**
BL21(DE3) (BBa_K567018, control plasmid)	8.78±0.22	18.50±0.29	**0**	**1.0**

Figura 3 (tese). Supplementary Figure S1. Schematic of the cloning procedure in the pOPT family of plasmids

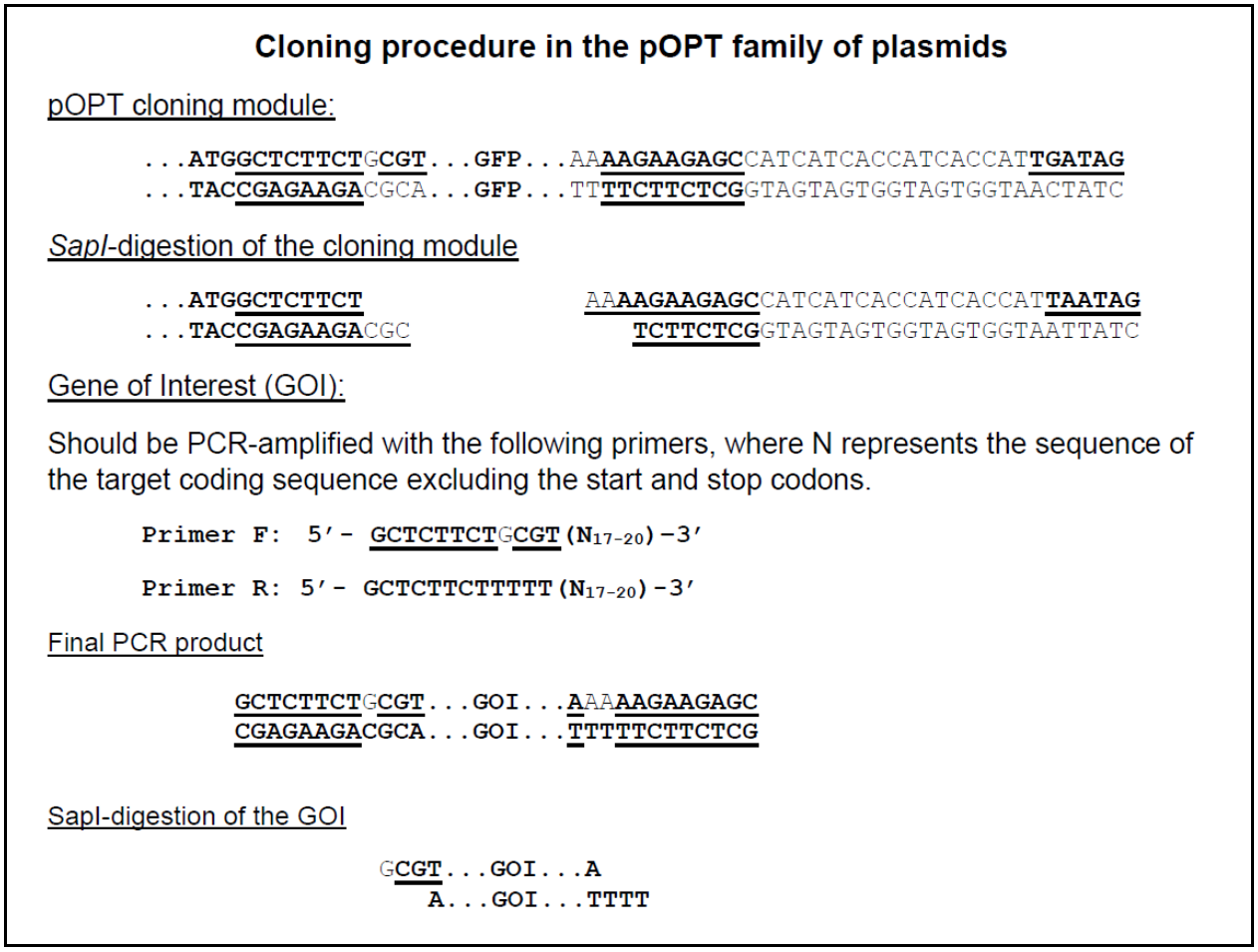


Figura 4 (tese). Supplementary Figure S2

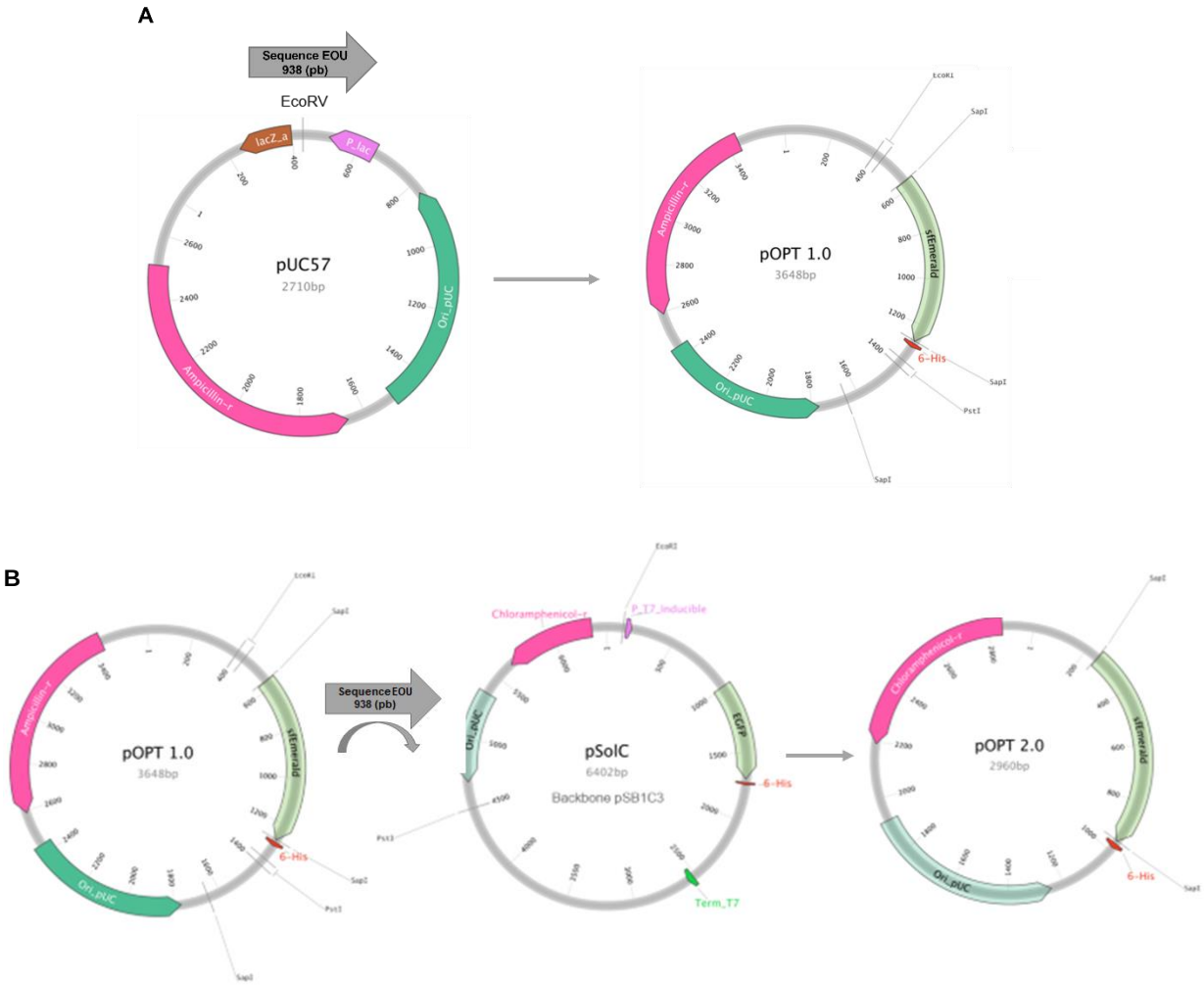


Fig. S2. Map of the pOPT plasmids and procedure for constructing the pOPT 2.0 vector. (A) Cloning of the expression operating unity (EOU) in the pUC57 vector giving rise to the pOPT 1.0 expression vector. (B)Transfer of the expression operating unity (EOU) from pOPT1.0 to the pSB1C3 backbone, using *EcoRI* and *PstI* digestion, giving rise to the pOPT 2.0 expression vector.

Figura 5 (tese). Supplementary Figure S3

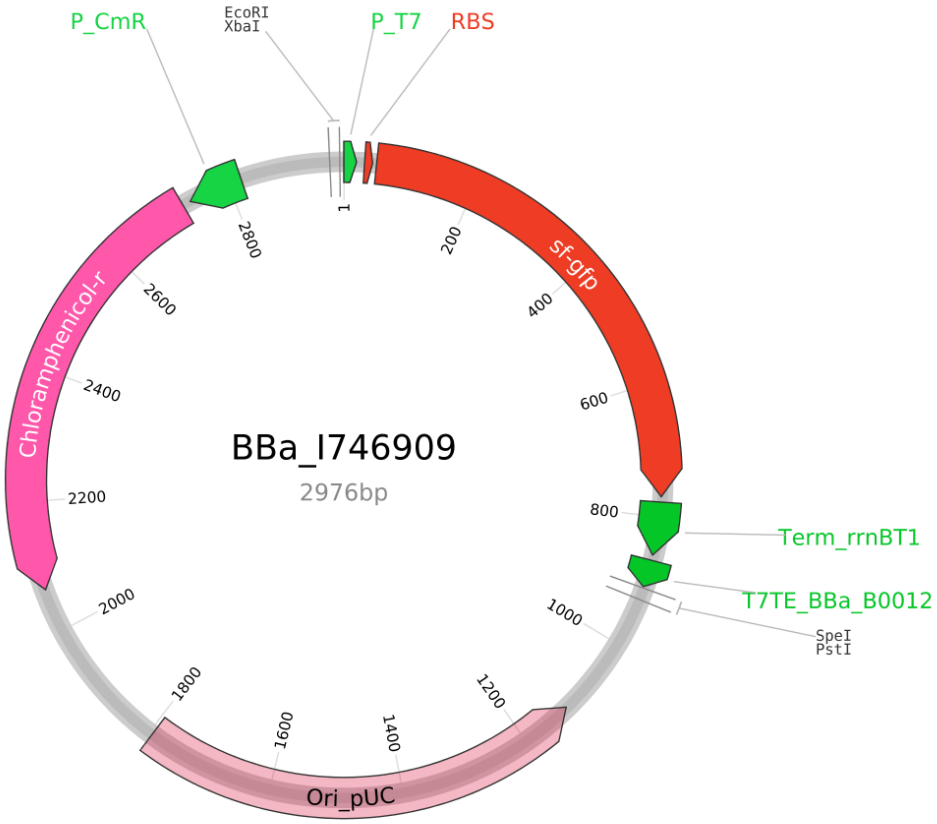


Fig. S3. Vector map of the plasmid BBa_I746909, used as a positive control in this study. Standard genetic elements that compose the plasmid BBa_I746909: wild-type T7 promoter, strong RBS (BBa_B0034) and two transcriptional terminators (B0010 and B0012).

Figura 6 (tese). Supplementary Figure S4

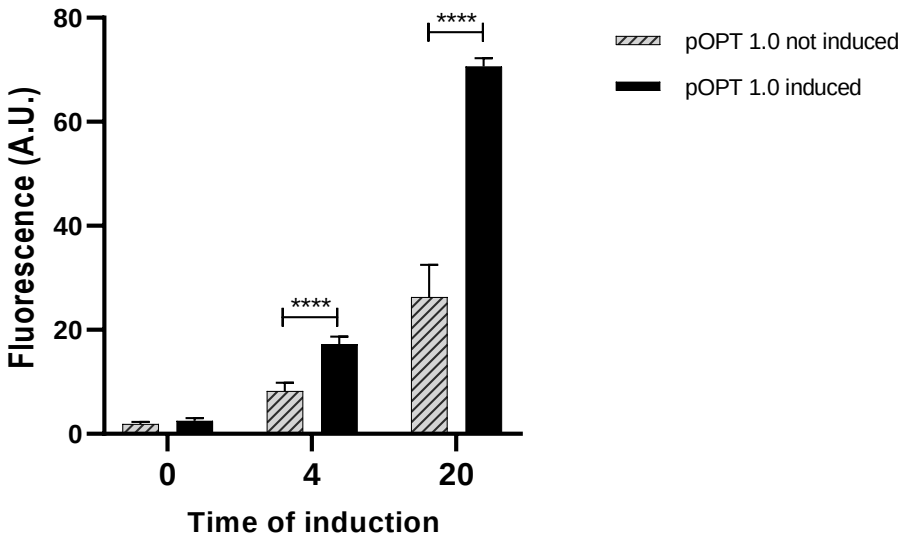


Fig. S4. sfGFP fluorescence emitted from pOPT1.0 transformed *E. coli* BL21(DE3), with and without induction by IPTG. Data are presented as mean ± s.d. (n =5). A statistically significant difference of $p < 0.0001$ relative to pOPT 1.0 (Bonferroni's multiple comparisons test) is denoted by ****.

Figura 7 (tese). Supplementary Figure S5

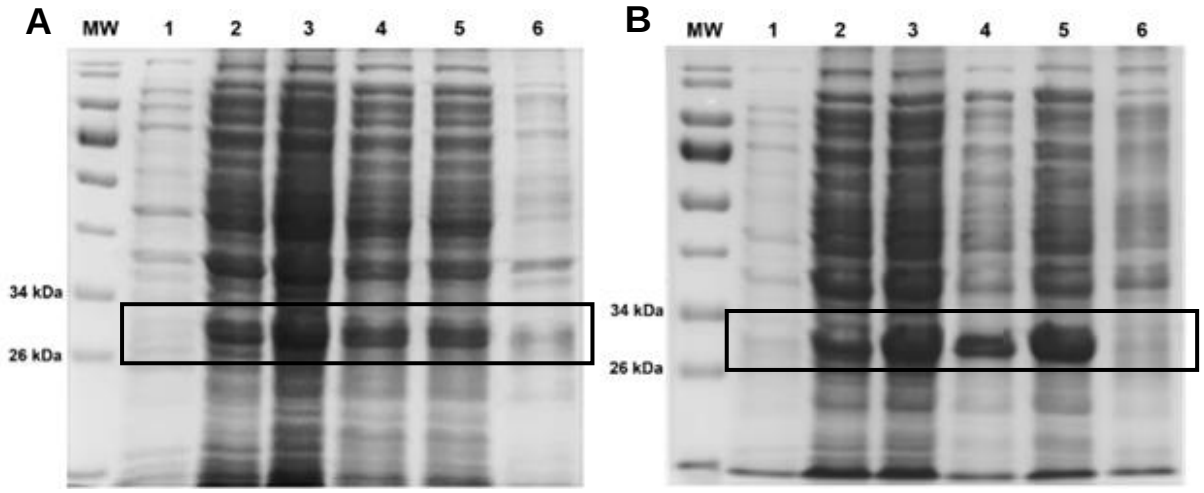


Fig. S5. Comparison of sfGFP protein expression in the BL21(DE3) strain transformed with the positive control plasmid BBa_I746909 or the pOPT 1.0 plasmid. (A) Expression of sfGFP protein in BL21(DE3) transformed with positive control plasmid BBa_I746909, analyzed on 12% SDS-PAGE. MW: Molecular weight standard (Thermo Scientific); 1: T0h; 2: T4h; 3: T20h; 4: Soluble Fraction 1*; 5: Soluble Fraction 2*; 6: Insoluble Fraction. (B) Expression of sfGFP protein in BL21(DE3) transformed with pOPT 1.0 plasmid, analyzed on 12% SDS-PAGE. MW: Molecular weight standard (Thermo Scientific); 1: T0h; 2: T4h; 3: T20h; 4: Soluble Fraction 1*; 5: Soluble Fraction 2*; 6: Insoluble Fraction. sfGFP expression is indicated by rectangles.
*Soluble Fraction 1: first resuspension of the cell lysate pellet with phosphate buffer.
*Soluble Fraction 2: the second resuspension of the remaining cell lysate pellet with a new volume of phosphate buffer to permit the recovery of the entire fraction of soluble proteins.

4 CAPÍTULO 2: CONSTRUÇÃO E CARACTERIZAÇÃO DE NOVAS MOLÉCULAS RECOMBINANTES HIPOALERGÊNICAS COM POTENCIAL APLICAÇÃO IMUNOTERAPÊUTICA

A tecnologia do DNA recombinante surgiu nos anos 70 e atualmente é fundamental para o desenvolvimento da biotecnologia, sendo uma das suas principais aplicações a produção de biomoléculas terapêuticas (BHANDARI; LIM; GARDNER, 2021; TUNGEKAR; CASTILLO-CORUJO; RUDDOCK, 2021). Dentre essas moléculas destaca-se a produção de proteínas recombinantes para compor formulação de vacinas para doenças respiratórias alérgicas, as quais ocorrem em todo o mundo e suas prevalências vem aumentado constantemente nas últimas décadas (ANSOTEGUI et al., 2020; KHALAF et al., 2022; MEGHJI et al., 2021; VALENTA; CAMPANA; NIEDERBERGER, 2017).

No contexto das doenças alérgicas temos que diferentes estímulos ambientais são capazes de desencadear processos alérgicos, sendo o principal fator de risco para esses processos a sensibilização aos alérgenos dos ácaros da poeira doméstica (ACEVEDO; ZAKZUK; CARABALLO, 2019; CALDERÓN et al., 2015; CARVALHO et al., 2013; MILLER, 2019). *Dermatophagoides pteronyssinus* é uma das principais fontes de alérgenos de ácaros da poeira e encontra-se distribuído em várias regiões do mundo (ACEVEDO; ZAKZUK; CARABALLO, 2019; LIMÃO et al., 2020; SADE; ROITMAN; KIVITY, 2010). Neste cenário, o alérgeno Der p 21 foi caracterizado como um novo importante alérgeno do ácaro *D. pteronyssinus*, apresentando alta capacidade de ligação à IgE e alta atividade alergênica (PULSAWAT et al., 2014; WEGHOFER et al., 2008). Sendo, portanto, considerado um potencial candidato não apenas para o diagnóstico de alergia mediada por IgE, mas também para imunoterapia alérgeno-específica (AIT) contra alergia causada por *D. pteronyssinus*.

A AIT é um tratamento para pacientes que sofrem de alergia associada à IgE, alternativo à farmacoterapia para reduzir os sintomas alérgicos (DOROFEEVA et al., 2021). Essa abordagem AIT aproveita o próprio sistema imunológico do paciente e é o único tratamento que modifica o curso natural da doença e previne a progressão de sintomas graves (CAMPANA et al., 2019; VALENTA; CAMPANA; NIEDERBERGER,

2017; ZHERNOV et al., 2019). A AIT tradicional baseia-se na utilização de extratos de alérgenos (SILVA et al., 2016; WALLNER; PICHLER; FERREIRA, 2013; YU; LIAO; TSAI, 2015). Porém, esse método apresenta várias desvantagens, o qual requer a administração de doses múltiplas e elevadas, podendo causar efeitos secundários graves e sistêmicos, levando até à morte (FERREIRA; WOLF; WALLNER, 2014; SILVA et al., 2016, 2020; VALENTA; CAMPANA; NIEDERBERGER, 2017).

Dessa forma, esforços estão sendo voltados para a produção de preparações de alérgenos bem padronizadas e eficazes por meio da tecnologia do DNA recombinante (AKINFENWA et al., 2021; SILVA et al., 2020; WALLNER; PICHLER; FERREIRA, 2013). No entanto, a imunoterapia com alérgenos recombinantes purificados também pode provocar efeitos secundários, variando de leve e local para sintomas mais graves e sistêmicos, com risco de vida (SILVA et al., 2016; VRTALA et al., 2004). Logo, para superar esses efeitos indesejáveis e para aumentar a eficácia, ferramentas modernas de engenharia de proteínas têm sido utilizadas em conjunto com novas ferramentas de bioinformática com o objetivo de desenvolver alérgenos hipoalergênicos (DA SILVA et al., 2020; THALHAMER et al., 2010; VALENTA; CAMPANA; NIEDERBERGER, 2017). Essas novas metodologias permitem a avaliação das características imunológicas dos alérgenos naturais e redução da sua alergenicidade (THALHAMER et al., 2010; VRTALA et al., 2004). Com essas ferramentas é possível realizar mudanças nas regiões de epítopos de células B, e manutenção das propriedades estimulantes de células T, como estratégia para formular imunoterapias hipoalergênicas e vacinas seguras contra alérgenos ambientais comuns (BERNARDINI et al., 2005; MACHADO et al., 2016; SILVA et al., 2016; SIRCAR et al., 2016; THALHAMER et al., 2010).

Considerando a importância da produção de proteínas hipoalergênicas, neste capítulo será apresentado um estudo que busca a construção e caracterização de novas moléculas recombinantes hipoalergênicas com potencial aplicação para imunoterapia de doenças alérgicas causadas pelo ácaro *D. pteronyssinus.* Será demonstrada a produção de moléculas recombinantes hipoalergênicas derivadas do alérgeno Der p 21 tipo-selvagem (wt), as quais foram projetadas utilizando um *pipeline* de análise de bioinformática estrutural, seguindo uma abordagem de mutagênese *in*

silico. Isso permitiu um desenho racional das variantes hipoalergênicas, o que tornou o processo de produção mais rápido e eficiente.

O *pipeline* de bioinformática estrutural e o desenho *in silico* das variantes mutantes hipoalergênicas do alérgeno Der p 21 aborda as seguintes etapas: recuperação de sequências e análises de domínios de proteínas e modificações pós-traducionais; predição da estrutura terciária da proteína, refinamento e validação; predição de epítopos de células B; predições de sítios de clivagem de epítopos de células T; mutagênese *in silico* do Der p 21; simulação de dinâmica molecular e análise de simulação imunológica.

Após construídos, os genes sintéticos que codificam o rDer p 21 wt e as quatro variantes mutantes foram clonados no vetor de expressão pOPT 2.0, recém-criado pelo nosso grupo de pesquisa (SANTOS et al., 2022a), para então serem expressos utilizando *E. coli* BL21(DE3) como hospedeiro. Por último, as novas variantes hipoalergênicas tiveram a sua alergenicidade avaliadas, por meio da investigação da capacidade de ligação à anticorpos IgE. Além disso, foram feitas análises *in silico* para avaliação da imunogenicidade.

4.1 ARTIGO 2: RATIONALLY DESIGNED HYPOALLERGENIC MUTANT VARIANTS OF THE HOUSE DUST MITE ALLERGEN DER P 21

Artigo publicado: *Biochim Biophys Acta Gen Subj.* (2022) 1866(4):130096. doi: 10.1016/j.bbagen.2022.130096.

Rationally designed hypoallergenic mutant variants of the house dust mite allergen Der p 21

Sara P. O. Santos[a,b,1], Ayrton B. P. Lisboa[a,c,1], Filipe S. R. Silva[a,b], Sandeep Tiwari[c], Vasco Azevedo[c], Álvaro A. Cruz[d], Eduardo S. Silva[a,e], Carina S. Pinheiro[a,b,e], Neuza M. Alcantara-Neves[a,b,e], Luis G. C. Pacheco[a,b,e,*]

[a] *Institute of Health Sciences, Federal University of Bahia, Salvador-BA, Brazil.*
[b] *Post-Graduate Program in Biotechnology, Federal University of Bahia, Salvador-BA, Brazil.*
[c] *Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte-MG, Brazil.*
[d] *ProAR Foundation and Federal University of Bahia, Salvador, Brazil*
[e] *Laboratory of Allergology and Acarology (LAA), Federal University of Bahia, Salvador-BA, Brazil.*

*** Corresponding author:** Prof. Luis G. C. Pacheco. Department of Biotechnology, Institute of Health Sciences, Federal University of Bahia, Salvador-BA, 40.110-100, Brazil. E-mail: luis.pacheco@ufba.br

[1] These authors contributed equally to this study.

Abstract

Background: Allergic diseases figure amongst the most common immune-mediated diseases worldwide, affecting more than 25% of the world's population. Allergic reactions can be triggered by house dust mite (HDM) allergens, of which the so-called group 21 of allergens is considered as clinically relevant.
Methods: Herein, we used a structural bioinformatics and immunoinformatics approach to design hypoallergenic mutant variants of the Der p 21 allergen of *Dermatophagoides pteronyssinus*, which were then recombinantly expressed in bacteria and tested for their IgE-reactivities. For this, we scanned the wild-type Der p 21 protein for all possible single amino acid substitutions in key IgE-binding regions that could render destabilization of the major epitope regions.
Results: Four main substitutions (D82P, K110G, E77G, and E87S) were selected to build mutant variants of the Der p 21 allergen, which were produced in their recombinant

forms; two of these variants showed reduced reactivity with IgE. Molecular dynamic simulations and immune simulations demonstrated the overall effects of these mutations on the structural stability of the Der p 21 allergen and on the profile of immune response induced through immunotherapy.
Conclusions: When produced in their recombinant forms, two of the Der p 21 mutant variants, namely proteins K110G and E87S, showed significantly reduced IgE reactivities against sera from HDM-allergic individuals ($n = 20$; $p<0.001$).
General significance: This study successfully translated a rational *in silico* mutagenesis design into low IgE-binding mutant variants of the allergen rDer p 21. These novel hypoallergens are promising to compose next-generation allergen-immunotherapy formulations in near future.

Keywords: Allergen-specific immunotherapy, *Dermatophagoides pteronyssinus*, Der p 21, recombinant hypoallergen, allergen engineering, immunoinformatics

1. Introduction

The most common manifestations of allergic diseases are immunoglobulin E (IgE)-mediated Type I hypersensitivity reactions (BLANK; HILGER, 2021; ZHERNOV et al., 2019), which affect a significant proportion of the global population (VALENTA; CAMPANA; NIEDERBERGER, 2017). One of the main risk factors to the development of Type I hypersensitivity is the sensitization to allergens from the house dust mites (HDM) (ACEVEDO; ZAKZUK; CARABALLO, 2019; CALDERÓN et al., 2015; CARVALHO et al., 2013; MILLER, 2019). *Dermatophagoides pteronyssinus and Dermatophagoides farinae* are the major sources of HDM allergens and they are widespread in many regions of the globe, triggering allergic symptoms (ACEVEDO; ZAKZUK; CARABALLO, 2019; LIMÃO et al., 2020; SADE; ROITMAN; KIVITY, 2010).

Among the *Dermatophagoides pteronyssinus* allergens, Der p 21 was characterized as a new important allergen, with a high capacity for binding to IgE antibodies and high allergenic activity (PULSAWAT et al., 2014; WEGHOFER et al., 2008). Pulsawat et al. 2014 suggested that Der p 21 can stimulate the production of IL-8 in airway epithelial cells through TLR2-dependent signaling, similarly to the mechanism described for Der p 2 (a major allergen of *D. pteronyssinus*) (PULSAWAT et al., 2014). The prevalence of reactivity to Der p 21 has also been investigated by different studies: 25% of subjects in a cohort of HDM-allergic patients in Thailand showed IgE reactivity to this allergen (PULSAWAT et al., 2014); in Austria, the reported reactivity was 26% (WEGHOFER et al., 2008); besides, similar percentages of IgE reactivity (*ca.* 20%) were also reported in studies from Singapore (KIDON et al., 2011) and Germany (POSA et al., 2017). The Der p 21 sequence has similarities with those of group 5 allergens, but does not show significantly high cross-reactivity. The sequence alignments have shown that Der p 21 and Blo t 5, from the mite *Blomia tropicalis*, have an amino acid (AA) sequence identity of *ca.* 40% (BESSOT; PAULI, 2011), while Der p 21 and Der p 5 have *ca.* 31% identity (GAO et al., 2007). Sequence studies also revealed that Der p 21 and Blo t 21 share *ca.* 41% of AA sequence identity (GAO et al., 2007). These features indicate that Der p 21 is a potential candidate not only for IgE-mediated allergy diagnosis, but also for allergen-specific immunotherapy (AIT) against allergy triggered by *D. pteronyssinus*.

AIT is a traditional treatment for patients suffering from IgE-associated allergy (DOROFEEVA et al., 2021) and can be more effective than pharmacotherapy and immunoreactive biologicals in reducing allergy symptoms. The AIT approach takes advantage of the patient's own immune system, and it is the only treatment that shows disease-modifying effects and prevents the progression of severe symptoms (CAMPANA et al., 2019; VALENTA; CAMPANA; NIEDERBERGER, 2017; ZHERNOV et al., 2019). However, traditional AIT based on crude mite extracts has important limitations due to the potential of inducing severe and systemic side effects, which require up-dose schedules and multiple administrations (SILVA et al., 2016, 2020; VALENTA; CAMPANA; NIEDERBERGER, 2017).

To circumvent these drawbacks, advancements in molecular allergen characterization led to the development of new forms of AIT based on recombinant purified proteins and hypoallergenic peptide derivatives (AKINFENWA et al., 2021; SILVA et al., 2020). In the hypoallergenization process, allergen derivatives can be produced by identifying the amino acid sequence determinants that are essential to IgE reactivity, and these regions are mutated or excluded at the genetic level, before recombinant protein production. Structural bioinformatics and immunoinformatics tools are generally employed to seek and rationally modify these protein regions in order to maintain the IgG reactivity, while decreasing IgE binding (DA SILVA et al., 2020; VALENTA; CAMPANA; NIEDERBERGER, 2017).

In this study, hypoallergenic recombinant molecules derived from the wild-type (wt) Der p 21 allergen were designed by using a dedicated structural bioinformatics analysis pipeline. Our approach involved *in silico* mutagenesis of the target allergen, allowing for the rational design of hypoallergenic variants. Prior knowledge of the allergen's three-dimensional structure and the location of B cell epitopes are required, enabling a targeted modification of these epitopes, but maintaining the immunogenicity of the allergen. We found that two mutant variants of Der p 21, namely K110G and E87S, presented significant decrease in IgE reactivities when tested with sera of HDM-allergic subjects, demonstrating their potential to compose hypoallergenic AIT formulations.

2. Materials and Methods

2.1. Structural bioinformatics and in silico mutagenesis of the Der p 21 allergen. The structural bioinformatics pipeline and the *in silico* design of the of Der p 21 hypoallergenic mutant variants are summarized in the Supplementary Material Fig.S1 and described in the subsequent items.

2.1.1. Sequence retrieval and analyses of protein domains and post-translational modifications. Database entries containing information about Der p 21 were searched in the Allergome relational database (http://www.allergome.org/) (MARI et al., 2009). Allergome entries contain links to the sequences in the UniProtKB protein database (http://www.uniprot.org/) (CONSORTIUM, 2019). Domain analyses were performed with the Hmmer Profile Search (http://hmmer.org/) against the pfam database (http://pfam.xfam.org/) (MISTRY et al., 2021). Domain prediction was used as a means to try to get further insights on the epitope positions within the Der p 21 allergen, *i.e.* whether or not major epitopes co-localize with predicted protein domains.

2.1.2. Protein tertiary structure prediction, refinement, and validation. Der p 21 wild-type sequences were submitted to the Robetta protein tertiary structure prediction server, using standard parameters (*de novo* modeling and homology modeling) (http://robetta.bakerlab.org/) (SONG et al., 2013). Refinement was carried out with FoldIt Standalone (KLEFFNER et al., 2017), which uses a RosettaCM algorithm to refine the protein sidechains and backbone based on free energy score (kcal/mol). Protein tertiary structure validation was then performed using the QMEANDisCo server (https://swissmodel.expasy.org/qmean/) (BENKERT; TOSATTO; SCHOMBURG, 2008) (Supplementary Table S1). 3D-structures visualization was performed with FoldIt and PyMOL *Molecular Graphics System* (https://pymol.org/).

2.1.3. Prediction of B-cell epitopes. Epitopes recognized by B cells were predicted on multiple prediction tools in order to build a consensus and a robust result. CBTope (http://osddlinux.osdd.net/raghava/cbtope/submit.php) (ANSARI; RAGHAVA, 2010) analysis were performed with a cut-off value for the SVM Threshold of -0,3 (default), because at this value, sensitivity and specificity were found equal during the development of the software. ElliPro (http://tools.iedb.org/ellipro/) (PONOMARENKO et

al., 2008) analyses were performed selecting the minimum score of 0.6 (default is 0.5) and maximum distance of 6 Angstroms (default). Both parameters guarantee good sensitivity and, particularly, good specificity. BCEPRED (http://crdd.osdd.net/raghava/bcepred/) (SAHA; RAGHAVA, 2004) analyses were performed using parameters of Antigenic Propensity by Kolaskar et al., 1990 (KOLASKAR; TONGAONKAR, 1990). DiscoTope (https://services.healthtech.dtu.dk/service.php?DiscoTope-2.0) (KRINGELUM et al., 2012) analyses were performed with the threshold for epitope prediction being set at -1.0. At this value, sensitivity was described as 0.30 and specificity as 0.85. Chimera software (https://www.cgl.ucsf.edu/chimera/) was used for the visualization of epitopes.

2.1.4. Predictions of T-cell epitope cleavage sites. MHC-II processed ligands prediction analyses were performed on the IEDB Tools MHCII-NP (http://tools.iedb.org/mhciinp/) (PAUL et al., 2018) server using sequences from Der p 21 wild-type and Der p 21 with all substitutions selected together. Comparison between Der p 21 and the mutants with 4 substitutions was performed to analyze the difference between the MHC II predicted peptides between the variants.

2.1.5. In silico mutagenesis of Der p 21. Amino acid residues with the highest scores on the B-cell epitopes prediction servers results were selected to build a consensus and undergo *in silico* mutagenesis. The positions 77 (glutamic acid), 82 (aspartic acid), 87 (glutamic acid), and 110 (lysine) of the Der p 21 wt allergen were selected to a following a scan for the most destabilizing mutations, performed with the MAESTROweb server (https://biwww.che.sbg.ac.at/maestro/web/) (LAIMER et al., 2015). A manual curation was performed to make sure that mutagenesis would not occur in sites that could lead to a complete disruption in the secondary structure or, consequently, in the tertiary structure of the original protein, such as, for example, proline in an alpha helix. The results of the *in silico* mutagenesis performed with MAESTROweb were then validated using the FoldIt Standalone program. Also, with the FoldIt *in silico* mutagenesis was performed using the '*mutate*' tool. After each substitution, a round of protein structure stabilization was carried in order to submit the protein structure to the effects of the aminoacidic change.

Finally, combinations were made between the most destabilizing mutations identified (E77G, D82P, E87S and K110G) for the production of Der p 21 wt variants: P1.D82P (Asp82Pro substitution), P2.K110G (Asp82Pro + Lys110Gly substitution), P3.E77G (Asp82Pro + Lys110Gly + Glu77Gly substitution) and P4.E87S (Asp82Pro + Lys110Gly + Glu77Gly + Glu87Ser substitution) (Table 1).

Tabela 5 (tese). Table 1. Predictions of the most destabilizing amino acid (AA) substitutions in the Der p 21 allergen and characteristics of the mutant derivatives constructed in this study.

		MAESTRO predictions[b]			FoldIt validation[b]	
AA position[a]	**Original AA**	**Suggested substitution**	***ΔΔG* (kcal/mol)**	**C*pred*[c]**	***ΔΔG* (kcal/mol)**	**Mutant derivatives** (AA substitutions)
82	Asp (D)	Pro (P)	1.275	0.904	11.641	**P1.D82P** (Asp82Pro)
110	Lys (K)	Gly (G)[d]	0.881	0.912	5.785	**P2.K110G** (Asp82Pro + Lys110Gly)
77	Glu (E)	Gly (G)	2.563	0.815	3.472	**P3.E77G** (Asp82Pro + Lys110Gly + Glu77Gly)
87	Glu (E)	Ser (S)	2.036	0.844	2.910	**P4.E87S** (Asp82Pro + Lys110Gly + Glu77Gly + Glu87Ser)

[a] Relative to Der p 21 wild-type sequences retrieved from Allergome DB.
[b] Starting Gibbs free energy value = -176.553 kcal/mol.
[c] Confidence estimation of the calculated *ΔΔG.* The values vary between 0.0 and 1.0, where 1.0 corresponds to a perfect prediction accuracy.
[d] Firstly predicted as Pro (P) but discarded due to potential alpha-helix break.

2.1.6. Molecular dynamics simulation. To evaluate the structural stabilities of the Der p 21 mutant variants in comparison to the wild-type allergen, we performed a molecular dynamics simulation experiment. Simulations were carried out by using the GROMACS 2020.3 package (ABRAHAM et al., 2015). CHARMM36 force field (HUANG; MACKERELL, 2013) was chosen to perform analysis. Proteins were solvated in a dodecahedron box with TIP3P water molecules. Na^+ and Cl^- ions were added to the system to achieve a neutral charged system. Electrostatic interactions were calculated using the Particle Mesh Ewald for long-range electrostatics method (CHEATHAM et al., 1995). V-rescale temperature coupling (modified Berendsen) was used to control

temperature (310 Kelvin). Berendsen pressure coupling (BERENDSEN et al., 1984) was used to control pressure (1.0 bar). The systems were subjected to unrestrained molecular simulation for 100 ns. Comparative analyses of structural deviations in wild-type and mutant structures were performed. RMSD and RMSF analyses were carried out using GROMACS tools. Graphs were plotted using MATLAB 2020 (MICHAEL HOLMBOE, 2020).

2.1.7. Immune simulation analysis. To predict the immune response induced by Derp21 (wt) or by the mutant proteins, we conducted an immune stimulation using C-ImmSim server (http://www.cbs.dtu.dk/services/C-ImmSim-10.1/) (RAPIN et al., 2010). For simulation, two injections containing 1,000 vaccine proteins each were given at intervals of one week. Time steps are set at 1 and 21 (each time step is 8 hours in real life, and time step 1 is the time of injection =0). The simulation steps have been modified to 1050, and the other parameters were kept as default.

2.2. Design and construction of recombinant plasmids. Table 2 describes all synthetic gene fragments and plasmids constructed in this study. Synthetic genes encoding rDer p 21 wt and the four mutant variants were obtained from Eurofins (Luxemburg), and then cloned using the pCR-Blunt II-TOPO vector (Invitrogen, Carlsbad, CA, EUA). The resulting plasmids were transformed into *E. coli* XL1Blue stain and recombinant clones were cultivated at 37°C, in Luria-Bertani (LB) broth, with kanamycin. Plasmids were recovered using the Nucleospin plasmid QuickPure kit (Macherey Nagel Co., Düren, Germany), the gene fragments were digested with *SapI* and then subcloned in the expression vector pOPT 2.0 (Table 2) (Supplementary Fig.S2), using standard protocols. Confirmation was performed by digestion with the restriction enzymes *EcoRI* and *PstI* (Promega, Madison, WI, USA) and analysis in a 1% agarose gel.

Tabela 6 (tese). Table 2. List of synthetic gene fragments, plasmids and bacterial strains used in this study.

	Characteristics	Source
Synthetic gene fragments		
*derp21*wt	387 bp; optimized synthetic gene fragment coding for the recombinant Der p 21 allergen.	This study
mut1	387 bp; optimized synthetic gene fragment coding for the Der p 21 mutant variant P1.D82P.	This study
mut2	387 bp; optimized synthetic gene fragment coding for the Der p 21 mutant variant P2.K110G.	This study
mut3	387 bp; optimized synthetic gene fragment coding for the Der p 21 mutant variant P3.E77G.	This study
mut4	387 bp; optimized synthetic gene fragment coding for the Der p 21 mutant variant P4.E87S.	This study
Strains		
E. coli XL1Blue	*endA1 gyrA96(nalR) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB+ laclq Δ(lacZ)M15] hsdR17(rK- mK+)*	Stratagene
E. coli BL21(DE3)	B F– ompT gal dcm lon hsdSB(rB–mB–) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS)	Stratagene
Plasmids		
pCR-Blunt II-TOPO	ColE1/ Kan^R, Neo^R; *lacZalpha*, *ccdB*	Invitrogen
pTP1:*derp21*wt	ColE1/ Kan^R, Neo^R; fragment encoding the recombinant Der p 21 allergen (rDer p 21), flanked by *SapI* sites.	This study
pTP2:*mut1*	ColE1/ Kan^R, Neo^R; cloned fragment encoding the Der p 21 mutant variant P1.D82P, flanked by *SapI* sites.	This study
pTP3:*mut2*	ColE1/ Kan^R, Neo^R; cloned fragment encoding the Der p 21 mutant variant P2.K110G, flanked by *SapI* sites.	This study
pTP4:*mut3*	ColE1/ Kan^R, Neo^R; cloned fragment encoding the Der p 21 mutant variant P3.E77G, flanked by *SapI* sites.	This study
pTP5:*mut4*	ColE1/ Kan^R, Neo^R; cloned fragment encoding the Der p 21 mutant variant P4.E87S, flanked by *SapI* sites.	This study
pOPT2.0	pUC ori/ Cm^R; Expression vector containing an optimized T7 promoter, strong RBS, fragment encoding GFP, 6x His-tag, T7 term. SapI-mediated scarless cloning.	This study
pOPT:*derp21*wt	pOPT2.0 containing a cloned gene fragment encoding the recombinant Der p 21 allergen (rDer p 21).	This study
pOPT:*mut1*	pOPT2.0 containing a cloned gene fragment encoding the Der p 21 mutant variant P1.D82P.	This study
pOPT:*mut2*	pOPT2.0 containing a cloned gene fragment encoding the Der p 21 mutant variant P2.K110G.	This study
pOPT:*mut3*	pOPT2.0 containing a cloned gene fragment encoding the Der p 21 mutant variant P3.E77G.	This study
pOPT:*mut4*	pOPT2.0 containing a cloned gene fragment encoding the Der p 21 mutant variant P4.E87S.	This study

2.3. Expression, purification, and analysis of recombinant Der p 21 and mutant variants. The expression vectors constructed in this study (Table 2) were transformed into chemically competent *E. coli* BL21(DE3) strains. The cells were grown at 37°C overnight in LB-agar plates supplemented with 34 µg.mL^{-1} of chloramphenicol. Confirmed clones were incubated in 10 mL LB broth (34 µg.mL^{-1} Cm) and induction of the target recombinant proteins was performed with 1 mM IPTG (Invitrogen, CA, California, EUA), at an O.D.$_{600nm}$ = 0.6. Cells were then incubated at 37°C for up to 20 hours. For recombinant protein purification, centrifuged bacterial pellets were collected at different time points during the induction protocol. Pellets were resuspended with 50 mM phosphate buffer, with or without 6 M urea (pH 8.0). Resuspended pellets were sonicated thrice, for 30 seconds at 50 Hz. His-tagged recombinant proteins were purified using the MagneHis™ Protein Purification system (Promega, Madison, WI, EUA), following manufacturer's instructions. Recombinant proteins were resolved in 12% SDS-PAGE and confirmed by Western blot using mouse anti-His IgG (GE Healthcare, 1:4000). Protein concentration was measured with the Qubit™ Protein Assay kit (Thermo Fisher Scientific, Waltham, MA, EUA).

2.4. Sera from HDM-allergic patients. In this study, two sets of sera were used. The first group of donors was composed of sensitized individuals, from both genders, aged ≥ 18 years, and living in the city of Salvador, Brazil. The second set of sera was obtained from individuals enrolled in Asthma and Allergic Rhinitis Control Program in Bahia (ProAR) (CRUZ et al., 2010). The allergy diagnosis, in both cases, was based on allergy symptoms (rhinitis, conjunctivitis, and/or asthma), positive skin prick test to *Dermatophagoides pteronyssinus*, and serum levels of IgE anti-*D. pteronyssinus* using immunoCAP (> 0,7 kUA/L). Sera from nonallergic individuals were used as control. The study was approved by the Ethics Committee on Research of the Faculty of Medicine of the Federal University of Bahia (CAAE 45376814.0.0000.5577). Blood samples were taken after informed consent was obtained from patients.

2.5. IgE reactivities of recombinant Der p 21 wt and of the mutant variants. Firstly, a *Dot blot* assay was performed to analyze the IgE reactivity to the Der p 21 wt and the mutant variants. For this, 2.5 µg of each recombinant protein was applied to a nitrocellulose membrane (GE Healthcare, Waukesha, WI, USA). Likewise, a crude

extract of *Dermatophagoides pteronyssinus* and a purified recombinant Der p 2 protein were used as positive controls. After 1.5h incubation with the antigens, membranes were blocked for 2 h at room temperature using a 3% skimmed milk and 5% tween-20 PBS solution. Then, membranes were incubated with pools of sera (1:5 diluted) at 4°C, for 16 h. A sera pool of five samples from sensitized patients was used, and another pool of five non-allergic patients was used as control. Following sera incubation, the membranes were incubated with a biotinylated anti-human IgE antibody (1:500, BD Biosciences, San Jose, CA) for 1 hour at room temperature and then incubated with Streptavidin-HRP (1:2500, BD Biosciences, San Jose, CA) for 45 min. The reacted bands were visualized using Enhanced Chemiluminescence (ECL) (GE Healthcare, Waukesha, WI, USA) and ImageQuant LAS 4000 (GE Healthcare, Waukesha, WI, USA). ImageJ software was also used to analyze the reacted spots.

The IgE reactivities of the different recombinant proteins were also evaluated by ELISA assays, according to a protocol previously developed by our group (CARVALHO et al., 2013). Briefly, high binding plates (*Greiner* Bio-One MICROLON™) were coated with 5.0 $\mu g.mL^{-1}$ of recombinant Der p 21 wt or the mutant variants, in a carbonate-bicarbonate buffer (pH 9.6–9.8), for 12h at 4°C. The plates were also blocked using a 10% BSA and 0.05% tween-20 PBS solution. The plate wells containing each Der p 21 variant was incubated with the positive or negative serum pool (diluted 1:5 in the blocking solution) for 16h at 4°C. The plates were then incubated with biotinylated anti-human IgE antibody (1:2000, BD Biosciences, San Jose, CA) for 1h at room temperature, followed by Streptavidin-HRP (1:1000, BD Biosciences, San Jose, CA) for 1h at room temperature. Finally, 0.1 mg.mL-1 TMB and 30% hydrogen peroxide were added to the plates for the reaction development. The optical density at 450 nm was measured by using a plate reader (Multiskan™ FC Microplate Photometer, Thermo Fisher Scientific, Vantaa, Finland).

For these ELISA assays, firstly, sera from 30 sensitized donors were used to access their individual IgE reactivities against a crude extract of *Dermatophagoides pteronyssinus* or rDer p 21. Following the confirmation of the donors with the highest reactivities to rDer p 21, the sera were again used to evaluate the IgE reactivities of the recombinant mutant variants.

2.6. Statistical analysis. GraphPad Prism software was used to plot the IgE reactivity data and to perform statistical analysis. Normality was verified with Shapiro–Wilk normality test and groups were analyzed by Wilcoxon test signed rank test. A value of $p < 0.05$ was considered as statistically significant.

3. Results and Discussion

Immunotherapy for HDM allergy using hypoallergenic allergens is currently considered as a promising alternative for more effective AIT (MARTÍNEZ et al., 2019; THALHAMER et al., 2010). To achieve this, recent studies have described protein engineering strategies to produce hypoallergenic mutant derivatives with reduced IgE binding, which retain epitopes recognized by T cells and induce blocking IgG antibodies (AGLAS et al., 2020; AKINFENWA et al., 2021; CURIN et al., 2018; DA SILVA et al., 2020; SILVA et al., 2016; THALHAMER et al., 2010). The overall aim of this strategy is to overcome the negative effects of allergen-specific immunotherapy, by maintaining the immunogenicity of AIT without increasing its allergenicity, then improving the risk-benefit ratio (AKDIS, 2012; LI et al., 2014).

In this study, we selected the Der p 21 allergen of *D. pteronyssinus* for hypoallergenization using the *in silico* mutagenesis approach, due to its high frequency of IgE recognition and allergenic activity (PULSAWAT et al., 2014; WEGHOFER et al., 2008). Previous studies have already shown that *in silico* designed hypoallergens can exhibit a significant reduction in IgE binding, while retaining T-cell stimulating properties (CHEN et al., 2014; LI et al., 2014; SIRCAR et al., 2016; THALHAMER et al., 2010). In addition to producing safer and more effective recombinant proteins for allergen immunotherapy, the *in silico* rational design of hypoallergens can also contribute to a reduction in the costs for obtaining the final product, given the reduction in experimental steps if compared to more traditional protein engineering methods such as gene shuffling (CHEN et al., 2014; DA SANTIAGO et al., 2012; LAIMER et al., 2015; ZHANG et al., 2012).

Firstly, we performed tridimensional structure modelling, refinement, and validation of the Der p 21 wild-type allergen (Fig. 1a; Supplementary Fig. S1). Following

predictions of linear and discontinuous B-cell epitopes (Supplementary Material Table S2 and Fig. S3), four amino acid residues were selected to undergo substitutions aimed at destabilization of the epitope regions of Der p 21, namely Glu(E)77, Asp(D)82, Glu(E)87, and Lys(K)110 (Fig. 1a-b). Three of these amino acid residues are present at the predicted conformational epitope E3 (Supplementary Fig. S4; Supplementary Table S2). Importantly, this region also co-localizes with an experimentally validated major conformational IgE-epitope of the Der p 21 allergen, termed peptide region P4 in the study by Curin and collaborators (CURIN et al., 2018). The computational screening for destabilizing mutations that affect B-cell epitopes was firstly used with success by Thalhamer and collaborators to generate hypoallergenic variants of the major pollen allergens of birch, Bet v 1, and timothy grass, Phl p 5 (THALHAMER et al., 2010). In our study, we have used the MAESTRO software (LAIMER et al., 2015) for the initial screening of the most destabilizing amino acid substitutions for each position, and then potential combinations of mutations were also evaluated with the aid of the FoldIt Standalone program (KLEFFNER et al., 2017). Basically, the MAESTRO tool uses machine learning to predict changes in the Gibbs free energy ($\Delta\Delta G$) of a given protein caused by single or combined amino acid substitutions. Additionally, it is possible to scan for the most stabilizing or destabilizing mutations. Positive values of $\Delta\Delta G$ indicate destabilization and MAESTRO also provides a confidence estimation for $\Delta\Delta G$ predictions (LAIMER et al., 2015).

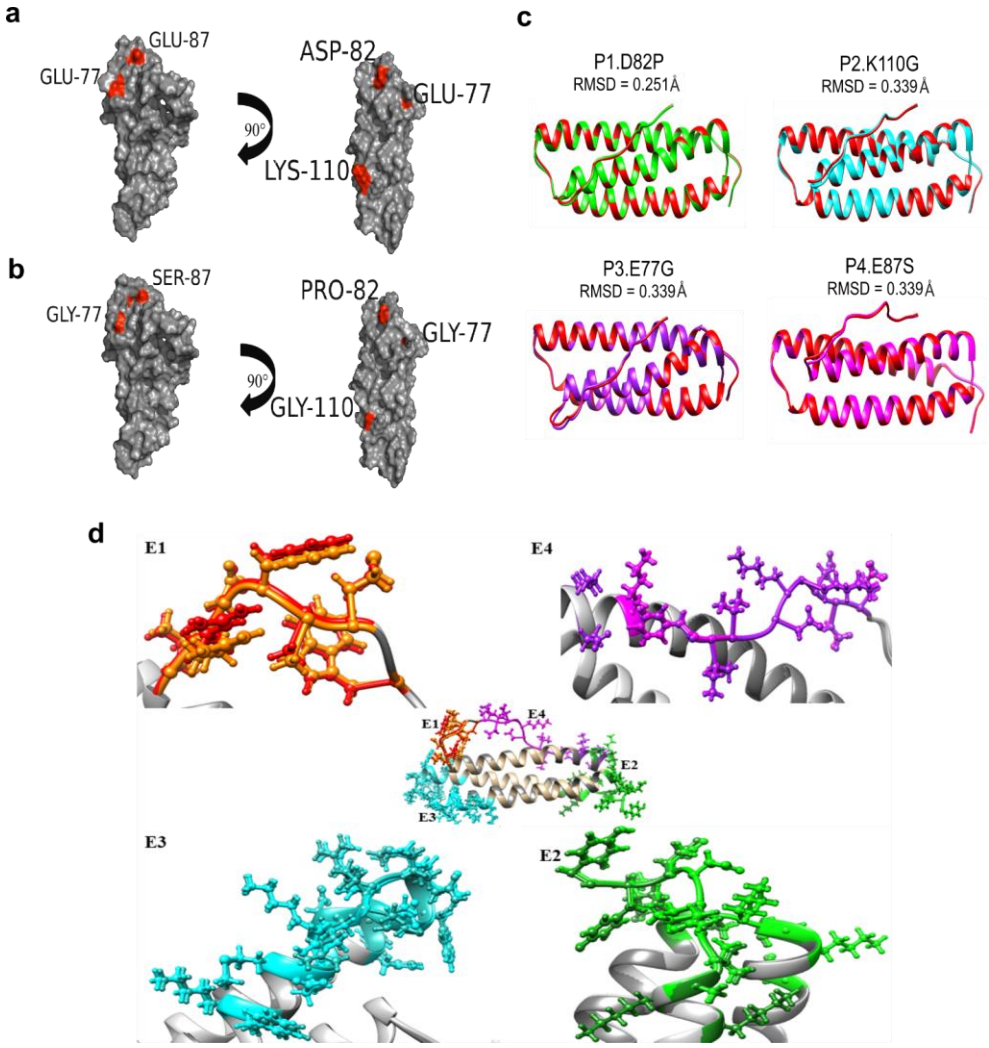


Figura 8 (tese). Fig. 1. *In silico* mutagenesis of Der p 21 allergen. (a) The four amino acid positions selected for mutagenesis. (b) Amino acid substitutions predicted as the most destabilizing in the conformational epitope regions. (c) Superimpositions of the 3D models of the four mutant variants with the Der p 21 wild-type allergen. Root-mean squared deviation (RMSD) values are indicated. (d) The four IgE-epitope regions (E1-E4) identified with high confidence in Der p 21. The zoomed in regions show superimpositions of the mutant variant containing the four substitutions (P4.E87S) with the Der p 21(wt).

The substitution of an aspartic acid residue at position 82 of the wild-type Der p 21 protein by a proline was selected by the MAESTRO software and then ranked by FoldIt as the most destabilizing mutation (Table 1); other changes that were also ranked as the top destabilizing substitutions were: change of a lysine at position 110 by a glycine; change of glutamic acid at position 77 by a glycine; and change of glutamic acid at position 87 by a serine (Table 1). These substitutions were combined to generate the

four mutant derivatives of Der p 21: P1.D82P, P2.K110G, P3.E77G, and P4.E87S (Fig. 1a-b; Table 1). Importantly, the selected mutations do not significantly alter the three helical bundle structure of the native Der p 21 allergen (Fig. 1c-d). A similar structure has been experimentally demonstrated for homologous proteins of the groups 21 and 5 of allergens (PANG et al., 2019). Interestingly, either drastic or non-drastic structural changes have been linked with the hypoallergenicity of a derivative for AIT (AGLAS et al., 2020; CHEN et al., 2012; DA SILVA et al., 2020; HOFER et al., 2017; WALLNER et al., 2011). In the case of some hypoallergenic hybrid to treat allergy caused by HDM, there was no drastic changes in comparison with the wild-type allergen (CHEN et al., 2012; DA SILVA et al., 2020).

In addition to evaluating the effect of the proposed substitutions on the structure of the epitope regions using the folded protein structures, we employed a molecular dynamics simulation (MDS) strategy to study the potential of destabilization of these mutations on the overall structure of the Der p 21 allergen in a dynamic environment. MDS is a tool that has been commonly adopted in computational protein engineering in recent years to predict the effects of newly introduced mutations on the structural stability of engineered proteins (CONDIC-JURKIC et al., 2018; GILL; MCCULLY, 2019; ZHANG; LAZIM, 2017). It is often employed with the aim of improving the stability of mutant derivatives of biotechnologically important proteins (GILL; MCCULLY, 2019; PIKKEMAAT et al., 2002; ZHANG; LAZIM, 2017). In our study, this strategy contributed to the detection of a significant structural instability introduced by substitution of the glutamic acid at position 77 of Der p 21 (Fig. 2). Results in Fig. 2a present dynamic simulations of 100 ns duration for each of the four mutant derivatives, in addition to the wild-type allergen. The backbone root mean squared deviation (RMSD) of the Der p 21 (wt) protein showed a rapid increase in the first 10 ns of the simulation and rapidly stabilized at around 4 Å (0.4 nm), with a gradual increase to *ca*. 6 Å over the course of the experiment. The RMSD of the mutant derivatives, in turn, showed large deviations from structural stability over the course of dynamic simulation (Fig.2a), even though all proteins had been classified as generally stable according to the instability indexes predicted by physicochemical parameters computed with the ProtParam tool (data not shown). It is noteworthy that allergens are generally considered as highly structured

proteins, when compared to non-allergenic proteins, and smaller RMSD fluctuations are expected in regions that correlate with major IgE epitopes (GARRIDO-ARANDIA et al., 2014). Our Root mean square fluctuations (RMSF) per residue results (Fig. 2b) clearly demonstrate that the specific amino acid substitutions performed in this study in the Der p 21 protein were responsible by large fluctuations in the epitope regions, with a significantly higher destabilizing effect attributed to the E77G mutation when introduced in the mutant protein 3 (P3.E77G) (Fig. 2b, merged panel). Interestingly, introduction of the fourth mutation (E87S) exerted a stabilizing effect on the mutant protein (Fig.2a and Fig.2b).

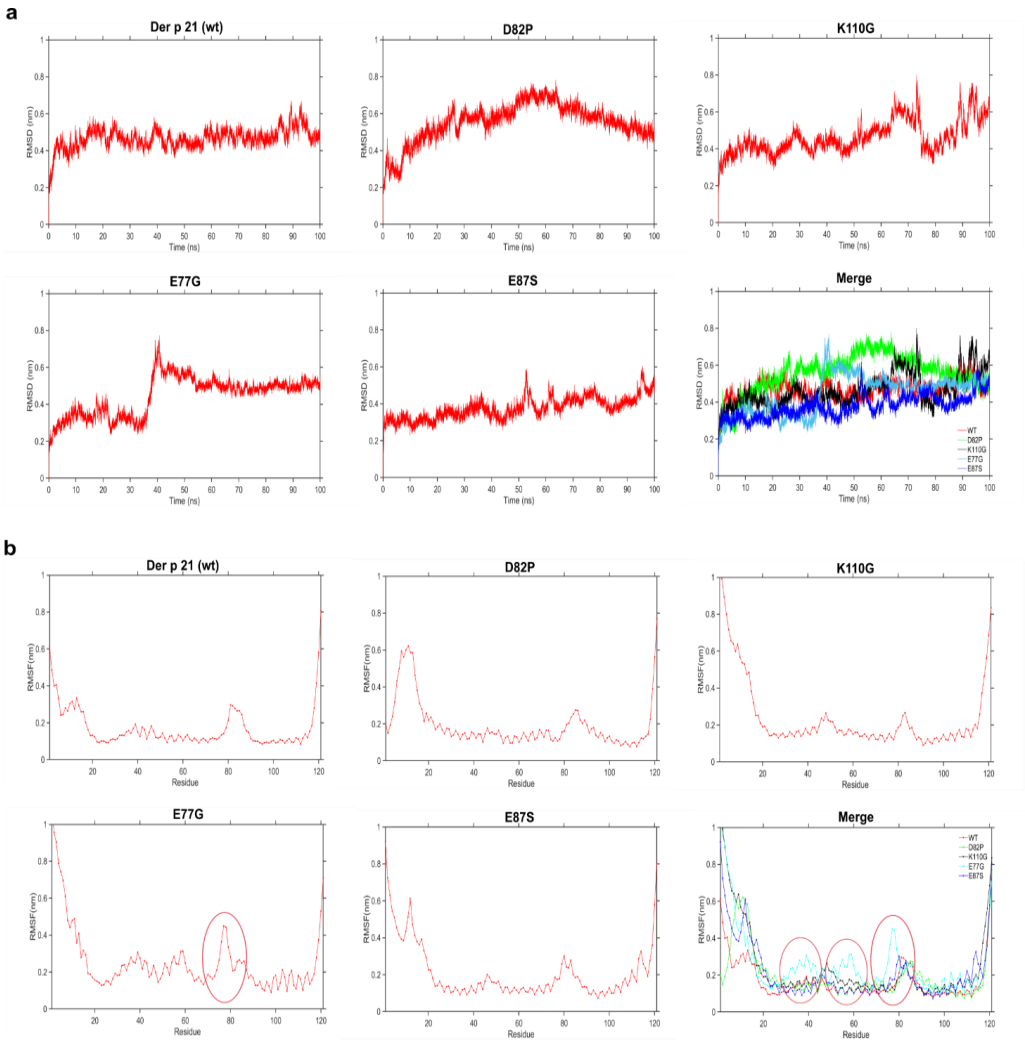


Figura 9 (tese). Fig. 2. Molecular dynamics simulation of the Der p 21 allergen and the four mutant variants. (a) RMSD values over a 100 ns simulation. (b) RMSF values per residue.

As aforementioned, an essential feature of a hypoallergenic protein, in order to be successfully used in next-generation allergy vaccines, is the retention of a helper T-cell reactivity comparable to that induced by the wild type allergen, despite the reduced IgE binding ability. In our study, a prediction of MHC class II bound T-cell epitopes of the Der p 21 wt allergen and the four mutant derivatives showed a conservation on cleavage sites following the four amino acid substitutions (Supplementary Table S3). Moreover, we used the C-ImmSim server to perform a dynamic immune simulation of the potential response induced by a prime-boost vaccination scheme (days 0 and 7) with the wild-type Der p 21 allergen and the four mutant variants (Fig. 3). Figures 3a and 3b show the predicted profiles of B-cells and T-helper cells induced by vaccination with the wild-type Der p 21 allergen, over a course of nearly a year. There is a predominantly induction of a T helper 1 (Th 1) profile of cells (Fig. 3; inset chart). Notably, a two-vaccinations regimen with the wild-type Der p 21 allergen is expected to induce high levels of INF-gamma, IL-12, TGF-beta, and IL-10 (Fig. 3c). No major alterations are predicted in the productions of cytokines and interleukins, when comparing the profiles induced by vaccination with the wild-type and the four mutant variants (Fig. 3d).

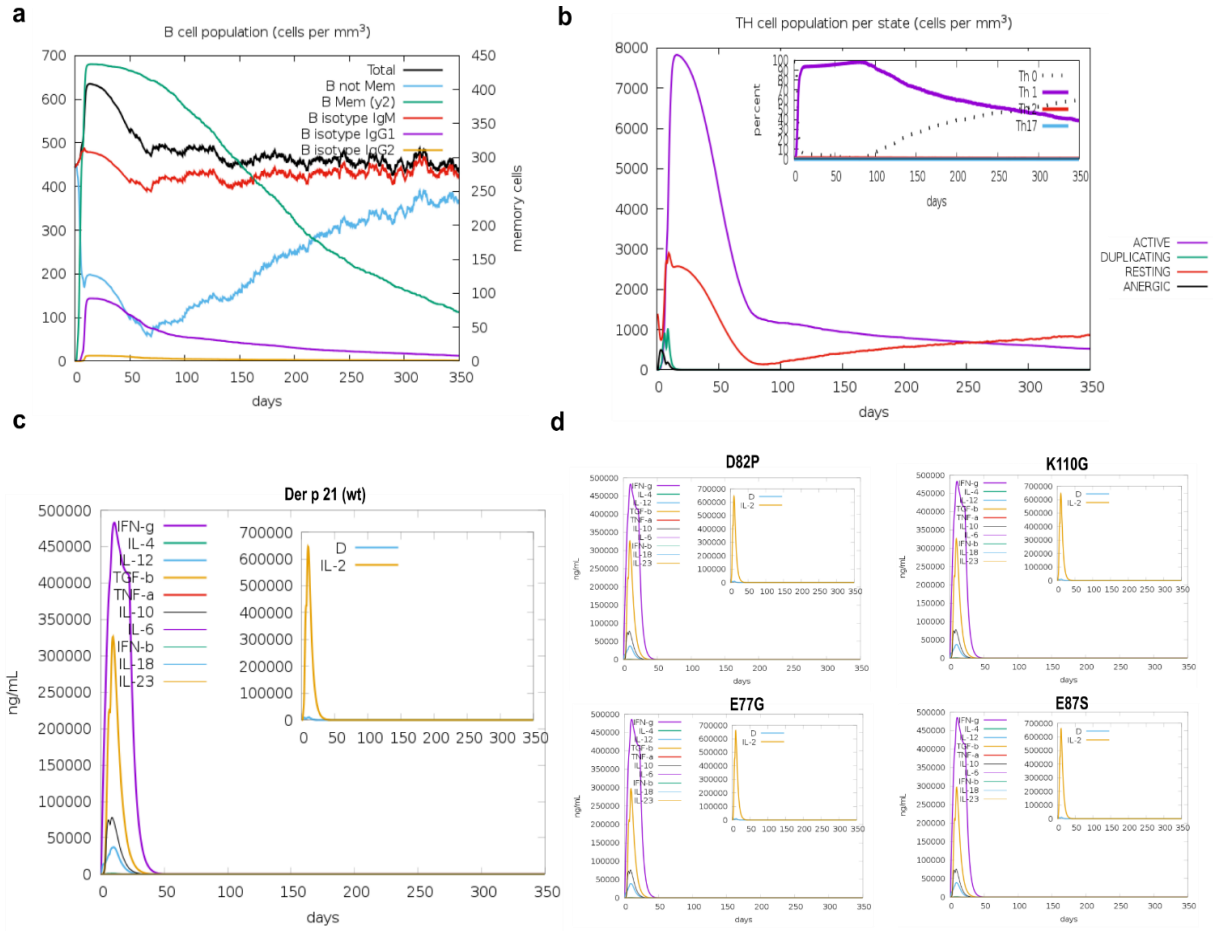


Figura 10 (tese). Fig. 3. Simulation of the immune response induced by vaccination with Der p 21 and the four mutant variants. (a) and (b) The profiles of B-cells and T-helper cells induced following a prime-boost vaccination regimen with the wild-type Der p 21 protein at days 0 and 7. The inset chart in (b) indicates the pattern of cellular immune response induced by vaccination. (c) Predicted cytokines and interleukins induced by vaccination with Der p 21(wt) over time. (d) Predicted cytokines and interleukins induced by vaccination with the mutant variants.

To evaluate whether the Der p 21 mutant variants designed by *in silico* mutagenesis would in fact present reduced IgE-binding activity when incubated with sera of allergic patients, we firstly implemented an effort to optimize the production and purification of the recombinant forms of these proteins expressed in bacteria. Recombinant Der p 21 (rDer p 21) had already been produced and recovered in its soluble form, without fusion to solubility or purification tags, by previous studies using *Escherichia coli* and the yeast *Pichia pastoris* as expression hosts (PULSAWAT et al., 2014; WEGHOFER et al., 2008). In this study, we chose to introduce a 6x-His

purification tag at the C-terminus of the five recombinant proteins, through translational fusion following scarless in-frame cloning at the pOPT2.0 expression vector (Table 2; Fig. 4a.). Figure 4b shows a schematic representation of the expected rDer p 21 protein (15,846 Da) expressed from pOPT2.0, following induction by IPTG. The amino acid substitutions in the recombinant mutant variants are highlighted in the figure, and all have been predicted as soluble by analyses with the tool Protein-Sol (BHANDARI; GARDNER, 2021; HEBDITCH et al., 2017), with solubility indexes higher than 0.6. The solubility of recombinant proteins in physiological buffers is critical for efficient and reliable production of therapeutic biomolecules, in order to overcome a well-known bottleneck in the biotechnology industry (CORREA; OPPEZZO, 2011; GARCÍA-FRUITÓS, 2014).

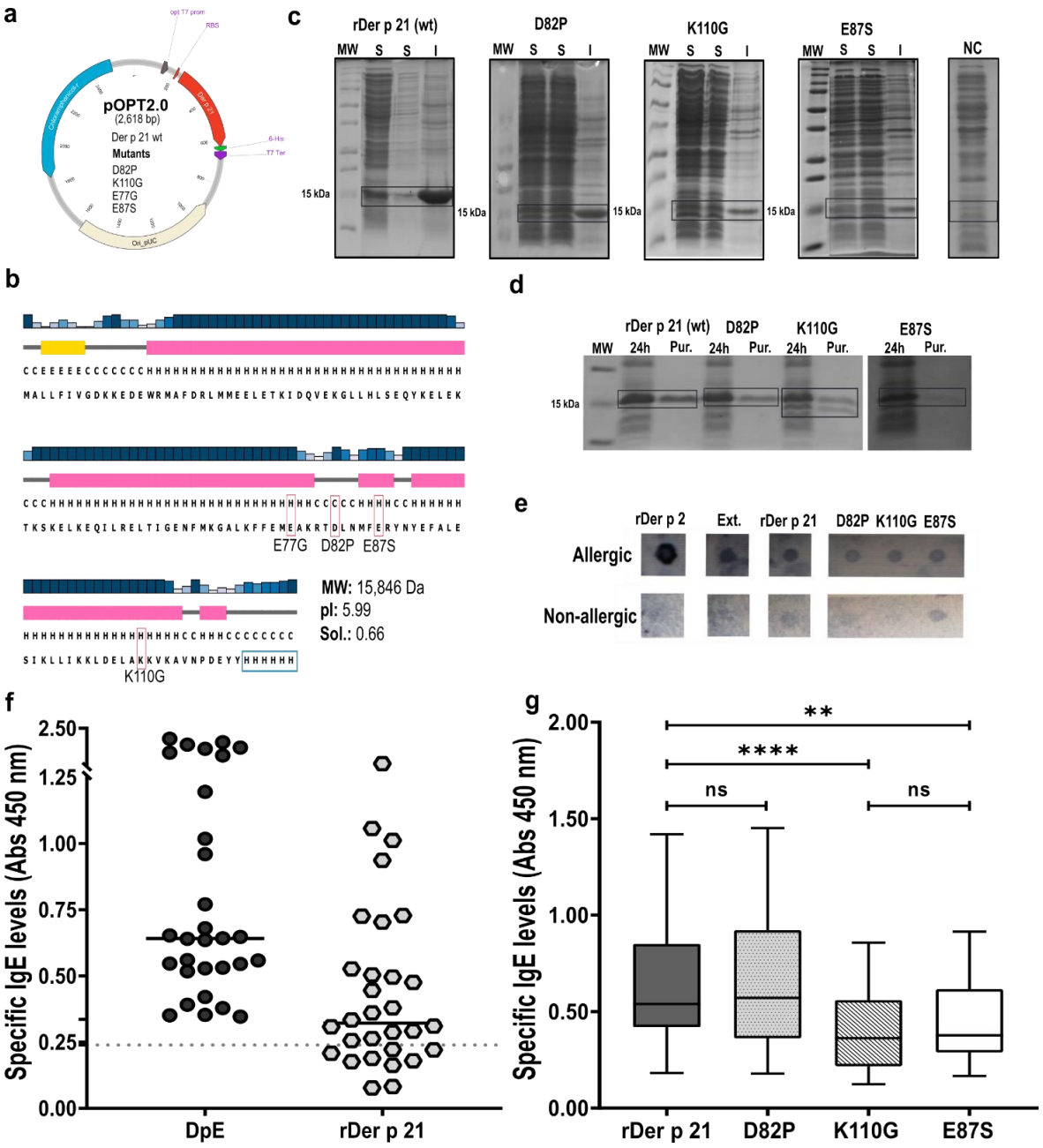


Figura 11 (tese). Fig. 4. Recombinant expression of the Der p 21 mutant variants and evaluation of IgE binding ability. (a) Schematic of the cloning of the gene fragments coding for rDer p 21 and the mutant variants in the pOPT2.0 expression vector. (b) Primary and predicted secondary structure of the recombinant proteins expressed in *E. coli*. The included 6x-His tag is indicated at the C-terminus. The pink regions indicate helical structures predicted by PSIPRED server. Darker blue bars indicate the confidence of secondary structure prediction per region. Molecular weight and isoelectric point of the rDer p 21 (wt) were calculated by ProtParam. Solubility score was calculated by Protein-Sol server. (c) 12% SDS-PAGE showing the expression profiles of rDer p 21 and the mutant variants. S = soluble fraction; I = insoluble fraction. NC = negative control, non-transformed BL21(DE3). (d) Purification of the recombinant proteins by affinity chromatography, following 24h induction. (e) Dot blot analysis with sera from 5 atopic individuals. (f) IgE reactivities of atopic sera (n = 30) against mite crude extract (DpE) or rDer p 21. (g) IgE reactivities of rDer p 21 and the mutant variants against highly reactive atopic sera (n = 20).

In fact, the successful recombinant *E. coli* clones expressed rDer p 21 and three of the mutant variants (P1.D82P, P2.K110G, and P4.E87S) in the soluble and insoluble fractions, as confirmed by 12% SDS-PAGE (Fig. 4c). These four recombinant proteins were purified to completion by affinity chromatography (Fig. 4d) with final protein yields ranging from 55.6 mg / L of culture to 60.0 mg / L of culture (Supplementary Table S4). Notably, the mutant protein 3 (P3.E77G), which was predicted as the most unstable by the dynamics simulation analysis, could not be successfully purified when expressed in *E. coli*, even when key expression parameters were modified, as suggested elsewhere (GUSTAFSSON et al., 2012; ROSANO; MORALES; CECCARELLI, 2019) (Supplementary Fig. S5). Corroborating previous studies that also produced rDer p 21 (PULSAWAT et al., 2014; WEGHOFER et al., 2008) or rDer f 21 (PANG et al., 2019), the ortholog protein from *Dermatophagoides farinae*, our Western blot results using an anti-His monoclonal antibody also suggest that the recombinant proteins may be found in solution partially as dimeric forms (Supplementary Fig. S6). Dimerization and aggregation can sometimes interfere in IgE-binding and hide epitopes in the allergen, leading to possible untrustworthy sensitization rates and usually reduction in reaction (NAJAFI et al., 2019; ZABORSKY et al., 2010). However, through dot blot analysis, we demonstrated that the wild-type rDer p 21 and all the three produced mutant variants are recognized by sera from allergic individuals (n = 5) (Fig. 4e), but react with different intensities. To quantify the differential levels of IgE binding, we then performed antigen-specific ELISA assays with these proteins, evaluating a set of sera from 30 *Dermatophagoides pteronyssinus*-sensitized patients. As shown in Fig. 4f, all donors reacted to the crude mite extract above the cut-off (0.3362), but only 20 of those reacted to rDer p 21. This IgE reactivity of 66.67% was higher than previous studies in both tropical and temperate climate regions of the globe (PULSAWAT et al., 2014; WEGHOFER et al., 2008) and classifies the allergen as a major allergen in Brazilian sensitized patients. Even though the allergen was partially dimerized in solution and considering the strong reaction in half of the rDer p 21-reactive individuals (Fig. 4f), this finding may lead us to believe this allergen may be included in the allergy diagnosis in the future, especially molecular based-diagnostic tools, an important trend in the field (ANSOTEGUI et al., 2020; PINHEIRO et al., 2021). Moreover, although previous studies

did find limited cross-reactivity between Der p 21 and Der p 5 (CURIN et al., 2018), studies on this topic for rBlo t 21 and rBlo t 5 cross-reaction with Der p 21 are still warranted. Considering the high inhibition behavior of Der f 21 against Blo t 5 and Blo t 21 IgE binding in another study (KIM et al., 2015), cross-reactivity between these two major *Blomia tropicalis* allergens and Der p 21 is a possibility that could directly interfere in allergy diagnosis.

The twenty reactive sera of allergic patients were again evaluated for rDer p 21 IgE reaction, but in comparison with IgE reaction against the mutant variants D82P, K110G, and E87S (Fig. 4g). There is a significant reduction in the IgE reactivities of the mutant proteins K110G and E87S, whereas the variant D82P, containing a single amino acid substitution, does not significantly differ from the wild-type recombinant allergen rDer p 21 (Fig. 4g). The reactivities of the proteins K110G and E87S are not significantly different, highlighting that they are similarly hypoallergenic. *The low IgE binding is a key feature in the production of an effective and safe hypoallergenic protein, suitable to be employed in AIT* (CASALE; STOKES, 2014; EICHHORN et al., 2019; ROBERTSON; TIRADO-RIVES; JORGENSEN, 2015; SILVA et al., 2016; THALHAMER et al., 2010; TSCHEPPE; BREITENEDER, 2017; VALENTA et al., 2011, 2016). The fact that the proteins containing combinations of amino acid substitutions (K110G and E87S) presented reduced IgE reactivities when compared to the single substitution in D82P corroborates similar studies with the Blo t 5 allergen of *Blomia tropicalis*, in which it was observed that the reduction in IgE binding was more significant in variants that contained multiple mutations (CHAN et al., 2008). In addition, our group has recently shown that mutation of four amino acid residues in a hybrid protein led to a hypoallergenic derivative with low IgE reactivity in sera from both allergic non-asthmatic and allergic asthmatic donors, including low avidity features and lack of basophils' activation (DA SILVA et al., 2020). Noteworthy, in the present study the inclusion of the fourth mutation E87S contributed to stabilize the overall structure of the recombinant protein, without changing the profiles of low IgE reactivity and the predicted T-cell activation; this might represent an advantage for further exploration of this mutant variant to compose a potential hypoallergenic vaccine formulation for house dust mite allergies.

Although further studies are still warranted, the present study successfully translated a rational *in silico* mutagenesis design into low IgE-binding mutant variants of the allergen rDer p 21. Additionally, we obtained high yields of recombinant expression and demonstrated, for the first time, that Der p 21 can be considered a major allergen in a tropical city of Brazil.

Author contributions

S.P.O.S., A.B.P.L., F.S.R.S., and S.T.: Investigation, Formal analysis, Methodology, Software, Data Curation, Visualization, Writing - Original Draft; F.S.R.S., E.S.S., and L.G.C.P: Investigation, Formal analysis, Methodology, Visualization, Writing - Review & Editing; V.A., A.A.C.: Resources, Project administration, Writing - Review & Editing; C.S.P., N.M.A.N., and L.G.C.P.: Conceptualization, Funding, Project administration, Supervision, Writing - Review & Editing.

Declaration of competing interest

S.P.O.S., A.B.P.L., F.S.R.S., E.S.S., C.S.P., N.M.A.N, and L.G.C.P. have filled a patent at Federal University of Bahia containing partial results of this project. It is currently pending at the National Institute of Industrial Property (INPI) of Brazil. The authors declare no other competing interests.

Acknowledgements

S.P.O.S. was recipient of a PhD scholarship from CAPES foundation, Brazil. V.A., A.A.C., N.M.A.N, and L.G.C.P. are recipients of research fellowships from CNPq-Brazil. This study was partially funded by FAPESB, through research grant #PNE0007/2014.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagen.2022.130096.

Supplementary Material

Rationally designed hypoallergenic mutant variants of the house dust mite allergen Der p 21

Sara P. O. Santos[a,b,1], Ayrton B. P. Lisboa[a,c,1], Filipe S. R. Silva[a,b], Sandeep Tiwari[c], Vasco Azevedo[c], Álvaro A. Cruz[d], Eduardo S. Silva[a,e], Carina S. Pinheiro[a,b,e], Neuza M. Alcantara-Neves[a,b,e], Luis G. C. Pacheco[a,b,e,*]

[a] *Institute of Health Sciences, Federal University of Bahia, Salvador-BA, Brazil.*
[b] *Post-Graduate Program in Biotechnology, Federal University of Bahia, Salvador-BA, Brazil.*
[c] *Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte-MG, Brazil.*
[d] *ProAR Foundation and Federal University of Bahia, Salvador, Brazil*
[e] *Laboratory of Allergology and Acarology (LAA), Federal University of Bahia, Salvador-BA, Brazil.*

*** Corresponding author:** Prof. Luis G. C. Pacheco. Department of Biotechnology, Institute of Health Sciences, Federal University of Bahia, Salvador-BA, 40.110-100, Brazil. E-mail: luis.pacheco@ufba.br

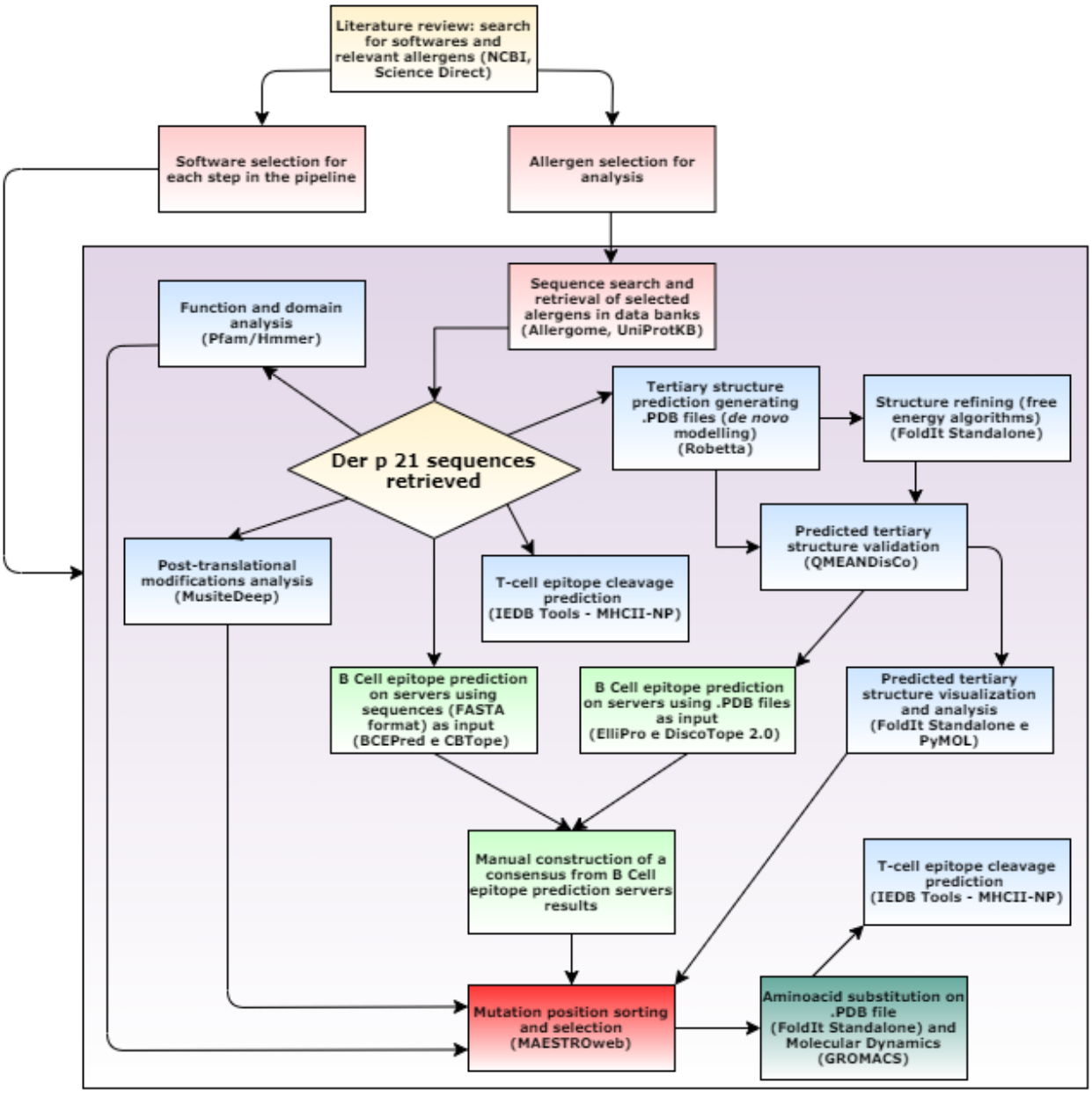


Figura 12 (tese). Supplementary Fig. S1. Overview of the bioinformatics pipeline used in this study for *in silico* mutagenesis of the Der p 21 allergen.

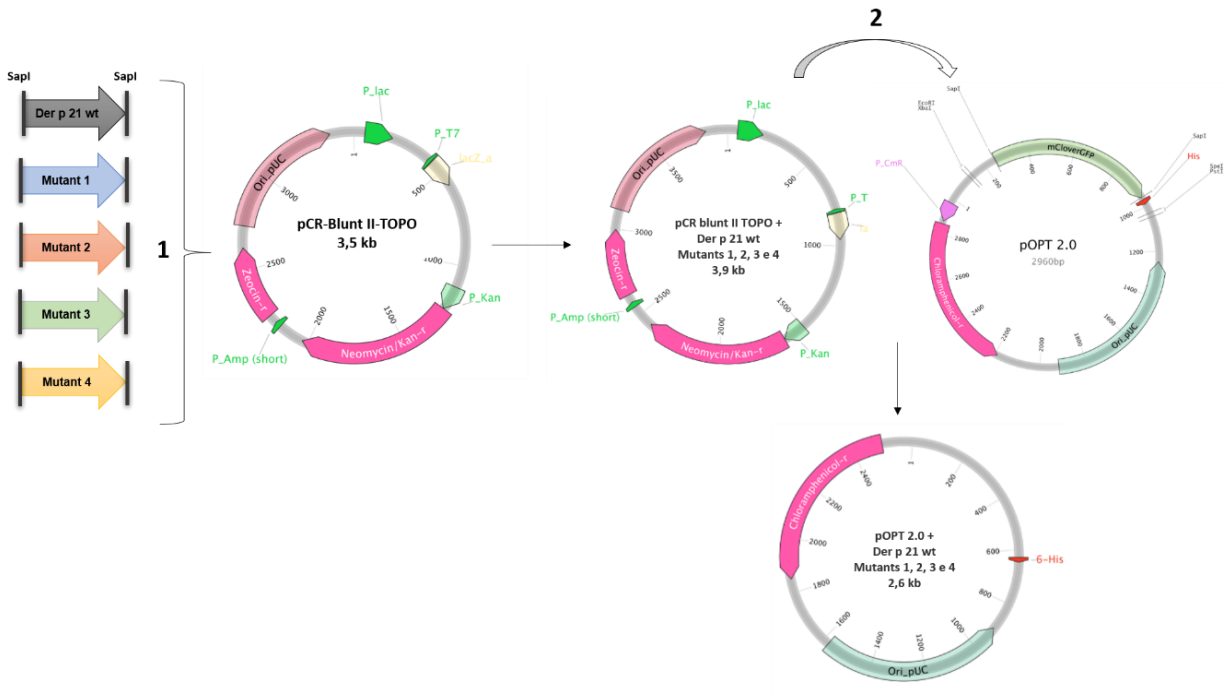


Figura 13 (tese). Supplementary Fig. S2. Overview of the cloning procedure to obtain the expression vectors containing the coding sequences for rDer p 21 and the four mutant variants.1) The synthetic gene fragments containing SapI restriction sites at the flanking regions were cloned in the pCR-Blunt II-TOPO vector. 2) Scarless sub-cloning of gene fragments in the pOPT 2.0 expression vector, using the type IIs restriction enzyme SapI, to generate a translational fusion with a C-terminal 6x-His tag.

B Cell epitope predictor	Epitopes predicted
CBTOPE	E GL LRELTIG E D FERY AKKVKAVN EY
ELLIPRO	FIVGDKKEDEWRMAFDRL MEAKRTDLNMFERY K KAVNPDEYY
BCEPRED	KKEDE Q QYKELEKTKSKELK I E D E K
DISCOTOPE	FIVGDKKEDEWRMAFDR E KE EKTKSKELKE R E RTDLNMFERY K VNPDEYY
Der p 21 sequence	**FIVGDKKEDEWRMAFDRLMMEELETKIDQVEKGLLHLSEQYKELEKTKSKELKEQILRELTIGENFMKGALKFFEMEAKRTDLNMFERYNYEFALESIKLLIKKLDELAKKVKAVNPDEYY**

Figura 14 (tese). Supplementary Fig.S3. Schematic representation of B cell epitopes of the Der p 21 allergen, predicted by different programs. A consensus was built analyzing the prediction scores and the frequency of prediction of a given amino acid. Amino acids highlighted in red were chosen for mutagenesis.

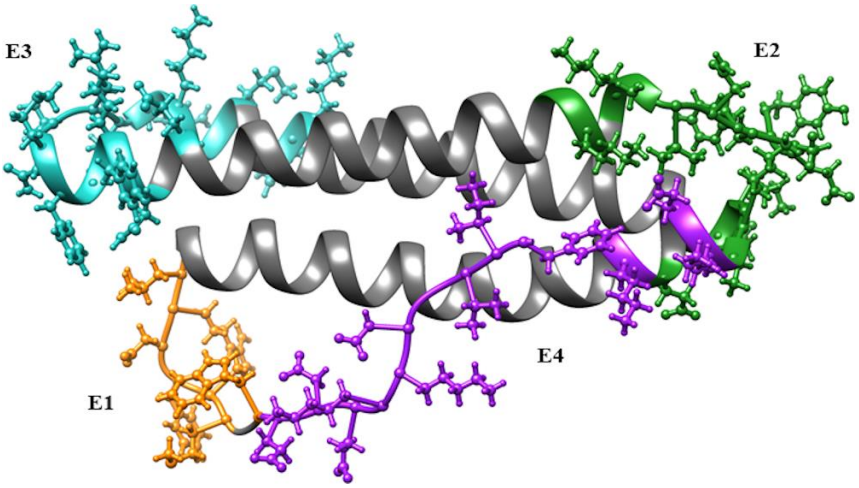


Figura 15 (tese). Supplementary Fig. S4: The Conformational B cell epitopes. Epitopes E1: Orange, E2: Forest Green, E3: Light sea Green and E4: Purple.

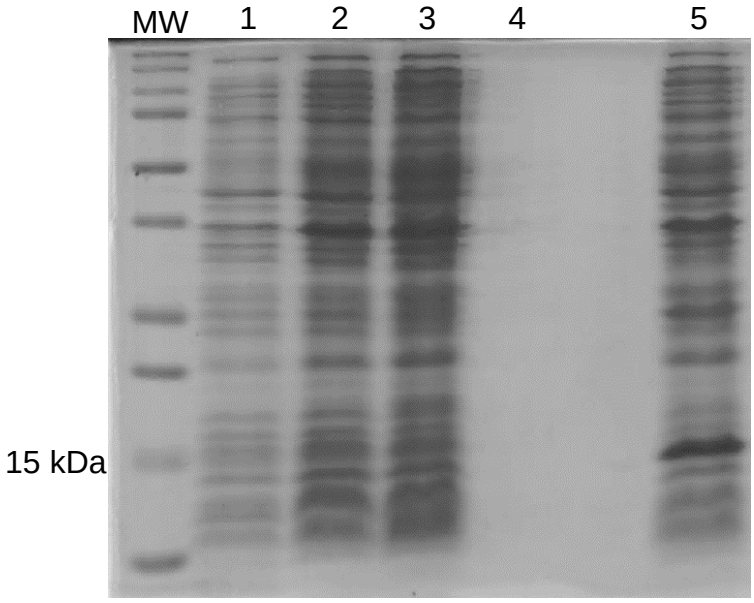


Figura 16 (tese). Supplementary Fig.S5: 12% SDS-PAGE showing a failed expression attempt of Der p 21 mutant variant protein 3 (E77G). MW = molecular weight marker Precision Plus Protein Standard (BioRad); Lane 1: 0 h; Lane 2: 4 h post induction; Lane 3: 24 h post induction; Lane 4: purified fraction; Lane 5: mutant protein 4 (E87S) at 24 hours post induction.

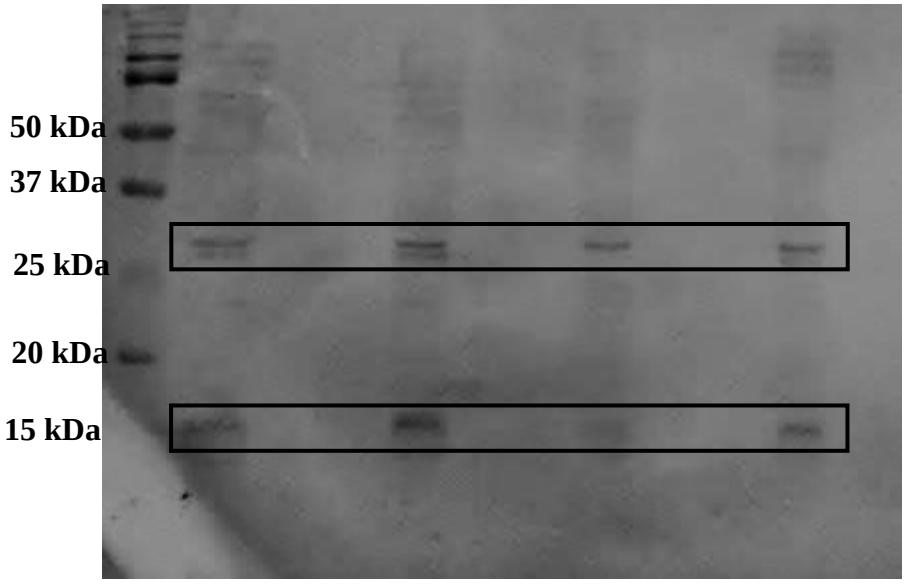


Figura 17 (tese). Supplementary Fig. S6: Western Blot of recombinant rDer p 21 (15.8 kDa) and its mutant variants expressed in BL21(DE3), using an anti-His antibody. MW: Molecular weight standard; Lane 1: rDer p 21; Lane 2: Mutant 1; Lane 3: mutant 2; Lane 4: Mutant 4. Positions expected for monomeric and dimeric forms are indicated.

Tabela 7 (tese). Supplementary Table S1. Qualitative Model Energy Analysis with Distance Contrainsts (QMEANDisCo) scores for each 3D protein structure model.

Protein	QMEANDisCo Global Score	QMEAN Error estimate
Der p 21 wt	0.69	± 0.08
D82P	0.67	± 0.08
K110G	0.66	± 0.08
E77G	0.65	± 0.08
E87S	0.65	± 0.08

Tabela 8 (tese). Supplementary Table S2: B cell conformational epitopes for Der p 21 (wt) and all four mutants.

Proteins	Epitopes	Residues	Number of Residues	Scores
Derp21 (wt)	**E1**	W11,M13,A14,F15,D16,R17,L18	7	0.815
	E2	E43,K46,T47,K48,S49,K50,E51,A114,V115, N116,P117,D118,E119,Y120,Y121	15	0.764
	E3	K72,F73,M76,E77,A78,K79,R80,T81,D82,L83,N84, M85,F86,E87,R88,Y89,N90,E92	18	0.764
	E4	F1,I2,V3,G4,D5,K6,K7,E8,D9, E10, K42	11	0.645
P1.D82P (mut1)	**E1**	W11,M13,A14,F15,D16,R17,L18	7	0.813
	E2	E43,K46,T47,K48,S49,K50,E51,A114,V115, N116,P117,D118,E119,Y120,Y121	15	0.765
	E3	K72,F73,M76,E77,A78,K79,R80,T81,P82,L83,N84, M85,F86,E87,R88,Y89,N90,E92	18	0.759
	E4	F1,I2,V3,G4,D5,K6,K7,E8,D9, E10, K42	11	0.648
P2.K110G (mut2)	**E1**	W11,M13,A14,F15,D16,R17,L18	7	0.824
	E2	E43,K46,T47,K48, S49,K50,E51,A114,V115, N116,P117,D118,E119,Y120,Y121	15	0.765
	E3	K72,F73,M76,E77,A78,K79,R80,T81,P82,L83,N84, M85,F86,E87,R88,Y89,N90,E92	18	0.712
	E4	F1,I2,V3,G4,D5,K6,K7,E8,D9, E10, K42,E45,K46	13	0.657
P3.E77G (mut3)	**E1**	W11,M13,A14,F15,D16,R17,L18	7	0.825
	E2	E43,K46,T47,K48,S49,K50,E51,A114,V115, N116,P117,D118,E119,Y120,Y121	17	0.716
	E3	K72,F73,M76,G77,A78,K79,R80,T81,P82,L83,N84, M85,F86,E87,R88,Y89,N90,E92	18	0.756
	E4	F1,I2,V3,G4,D5,K6,K7,E8,D9, E10, K42,E45,K46	13	0.66
P4.E87S (mut4)	**E1**	W11,M13,A14,F15,D16,R17,L18	7	0.824
	E2	E43,T47,K48,S49,K50,E51,G110,K111,K113,A114,V115, N116,P117,D118,E119,Y120,Y121	17	0.714
	E3	K72,F73,M76,G77,A78,K79,R80,T81,P82,L83,N84, M85,F86,S87,R88,Y89,N90,E92	18	0.758
	E4	F1,I2,V3,G4,D5,K6,K7,E8,D9, E10, K42,E45,K46	13	0.658

The Ellipro tool predicted 4 conformational B-cell epitopes with scores >0.6 for all of the proteins

Tabela 9 (tese). Supplementary Table S3: Predictions of MHC class II cleavage sites in Der p 21 and the protein with the four substitutions (P4.E87S).

Protein	Peptide start	Peptide end	Peptide length	Peptide	N motif	C motif	Cleavage probability score	Cleavage probability percentile rank
Der p 21 WT	27	40	14	IDQVEKGLLHLSEQ	KID	EQY	0.54528	0.0
	16	29	14	DRLMMEELETKIDQ	FDR	DQV	0.51974	0.09
	5	17	13	DKKEDEWRMAFDR	GDK	DRL	0.48917	0.18
	27	42	16	IDQVEKGLLHLSEQYK	KID	YKE	0.4846	0.27
	16	31	16	DRLMMEELETKIDQVE	FDR	VEK	0.44603	0.36
Der p 21 with 4 substitutions	81	94	14	TPLNMFSRYNYEFA	RTP	FAL	0.60079	0.0
	27	40	14	IDQVEKGLLHLSEQ	KID	EQY	0.54528	0.09
	16	29	14	DRLMMEELETKIDQ	FDR	DQV	0.51974	0.18
	5	17	13	DKKEDEWRMAFDR	GDK	DRL	0.48917	0.27
	27	42	16	IDQVEKGLLHLSEQYK	KID	YKE	0.4846	0.36

Tabela 10 (tese). Supplementary Table S4. Quantitative yields of purified rDer p 21 (wt) and mutant variants mut1 (D82P), mut2 (K110G), and mut4 (E87S).

Protein	Purification with native condition	Purification with denaturing condition
Der p 21 wt	41.6 mg/L	60.0 mg/L
Mutant 1	38.5 mg/L	59.2 mg/L
Mutant 2	33.0 mg/L	56.8 mg/L
Mutant 4	33.0 mg/L	55.6 mg/L

5 CONCLUSÕES E CONSIDERAÇÕES FINAIS

A biologia sintética se desenvolveu nas últimas duas décadas como um campo emergente de pesquisa, multiprofissional e multidisciplinar, proporcionando vários avanços para a biotecnologia, incluindo o desenvolvimento de vacinas, diagnóstico molecular e aplicações terapêuticas (BROOKS; ALPER, 2021; KHALIL; COLLINS, 2010; TAN et al., 2021). Essa área de pesquisa reúne biólogos e engenheiros para projetar e construir novos componentes biomoleculares, permitindo a otimização de sistemas biológicos (KAHL; ENDY, 2013; KHALIL; COLLINS, 2010).

O surgimento da biologia sintética aprimorou a produção de proteínas recombinantes, considerado um importante processo biotecnológico e de grande interesse industrial (DRAGOSITS; NICKLAS; TAGKOPOULOS, 2012; LOZANO TEROL et al., 2021; TUNGEKAR; CASTILLO-CORUJO; RUDDOCK, 2021). Isso porque o conhecimento dos fatores envolvidos na produção de proteínas recombinantes levou ao desenvolvimento de ferramentas de biologia sintética direcionadas à engenharia do dogma central da biologia molecular. Essas ferramentas melhoraram e expandiram a capacidade da maquinaria celular, contribuindo para a produção otimizada de proteínas (LOZANO TEROL et al., 2021; YOUNG; ALPER, 2010).

O avanço das técnicas de biologia sintética levou a redução do preço de fragmentos de DNA, e possibilitou uma revolução no surgimento de sequências sintéticas que não necessitam de um molde natural para serem sintetizadas. Com isso é possível editar os elementos genéticos envolvidos na síntese de proteínas e até criar dispositivos que não existem na natureza (YOUNG; ALPER, 2010). Vários trabalhos vêm sendo relatados, demonstrando que o uso da biologia sintética permite a criação de novas partes biológicas ou até mesmo de novos vetores de expressão, os quais levam ao aumento da produção de proteínas recombinantes em bactérias (FIRAT DUZENLI; OKAY, 2020; MIKIEWICZ et al., 2019; SANTOS et al., 2022a; ZHANG et al., 2021).

Atualmente a tentativa de aumentar o nível de expressão de proteínas recombinantes em sistemas bacterianos, bem como melhorar a qualidade da produção, tornou-se um aspecto altamente relevante, tendo em vista o crescente uso das proteínas recombinantes para diversas aplicações (KAUR; KUMAR; KAUR,

2018; LOZANO TEROL et al., 2021). Com o intuito de suprir essa necessidade foi criado pelo nosso grupo de pesquisa um vetor de expressão de forma racional, utilizando partes genéticas padronizadas para melhorar a produção de proteínas recombinantes em *Escherichia coli* (SANTOS et al., 2022a). Esse novo vetor gerou uma produção eficiente da proteína repórter *sf*GFP (unidades de fluorescência relativa/RFU = 70,62 ± 1,62 A U.), principalmente na forma solúvel, quando comparado a um vetor de expressão utilizado como controle (plasmídeo BBa_I746909; RFU = 59,68 ± 1,82 A U.), demonstrando a importância de reutilizar genes previamente caracterizados. Além disso, nesse estudo foi construída uma nova variante do promotor T7, a qual apresentou uma atividade transcricional melhorada quando comparada com o promotor T7 tipo-selvagem. Mutalik e colaboradores (2013a) também demonstraram a importância de se construir uma unidade de expressão racional, testando os elementos genéticos sintéticos e avaliando a influência que eles possuem na produção de proteínas recombinantes (MUTALIK et al., 2013a). Assim, a caracterização prévia dos elementos genéticos, permite que eles sejam utilizados em estudos posteriores de forma mais precisa e confiável, o que diminui os custos e o tempo gasto até a obtenção do produto final.

Esse novo vetor de expressão recém-criado também teve a sua funcionalidade avaliada com a produção de genes de interesse biotecnológico. A sua capacidade de produzir uma proteína de origem eucariótica de difícil expressão foi demonstrada pela produção do alérgeno Der p 21 do ácaro da poeira doméstica *Dermatophagoides pteronyssinus* e das suas variantes hipoalergênicas (SANTOS et al., 2022b). Em tentativas anteriores de expressar o alérgeno Der p 21 a partir de vetores comercialmente disponíveis, como pD444, não conseguimos uma expressão bem-sucedida em *E. coli*, mesmo quando diferentes condições de expressão foram testadas. Destaca-se que após a clonagem no plasmídeo pOPT 2.0, portador da nova UOE, tanto o alérgeno Der p 21 como seus derivados mutantes puderam ser produzidos em *E. coli* com rendimentos finais de proteína solúvel variando de 55,6 a 60 mg.L-1 de cultura.

No estudo citado acima, moléculas recombinantes hipoalergênicas derivadas do alérgeno Der p 21 wt foram projetadas usando um *pipeline* de análise de bioinformática estrutural. Foi realizada uma abordagem de mutagênese *in silico* do alérgeno alvo, permitindo o desenho racional das variantes hipoalergênicas, o que tornou o processo de produção mais rápido e eficiente. Para isso, é importante o

conhecimento prévio da estrutura tridimensional do alérgeno e da localização dos epítopos das células B, o que possibilita uma modificação direcionada desses epítopos, mas mantendo a imunogenicidade do alérgeno.

Os genes construídos foram clonados no vetor de expressão pOPT 2.0, e os ensaios de expressão foram realizados em cepas de *E. coli* BL21(DE3). Após os testes de caracterização das proteínas produzidas ficou comprovado que duas variantes mutantes do Der p 21 wt, K110G e E87S, apresentaram diminuição significativa nas reatividades de IgE quando testadas com soros de indivíduos alérgicos a ácaros da poeira doméstica, demonstrando seu potencial para compor formulações hipoalergênicas de AIT. Além disso, uma previsão de epítopos de células T ligados a MHC de classe II, do alérgeno Der p 21 wt e dos quatro derivados mutantes, mostraram uma conservação nos locais de clivagem após as quatro substituições de aminoácidos, o que sugere uma retenção da reatividade de células T comparável àquela induzida pelo alérgeno selvagem, uma característica essencial de uma proteína hipoalergênica. Por fim, temos que os resultados da reatividade de IgE ao Der p 21 wt revelaram, pela primeira vez, que esse alérgeno pode ser considerado um alérgeno importante em uma cidade tropical do Brasil.

Contudo, a criação de novos vetores de expressão otimizados é uma alternativa promissora para melhorar a produção de proteínas recombinantes de interesse biotecnológico em sistemas bacterianos. Com esse avanço várias áreas de pesquisa poderão ser beneficiadas, gerando um impacto significativo no futuro das aplicações da Biotecnologia moderna, incluindo o desenvolvimento de vacinas para doenças respiratórias alérgicas baseadas em proteínas recombinantes hipoalergênicas.

6 PERSPECTIVAS FUTURAS

- Utilizar o novo vetor de expressão recém-criado para produção de outras proteínas de interesse biotecnológico.
- Explorar a contagem de sensibilização ao alérgeno Der 21 para diagnóstico de doenças alérgicas.
- Realizar mais estudos de caracterização imunológica das variantes mutantes, 2 (K110G) e 4 (E87S), para comprovar que elas podem ser de fato utilizadas para compor formulações de vacinas de nova geração para doenças alérgicas causadas pelo ácaro *Dermatophagoides pteronyssinus.*

7 REFERÊNCIAS

ABBAFATI, C. et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. **The Lancet**, v. 396, n. 10258, p. 1204–1222, 2020.

ABRAHAM, M. J. et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. **SoftwareX**, v. 1–2, p. 19–25, 2015.

ACEVEDO, N.; ZAKZUK, J.; CARABALLO, L. House Dust Mite Allergy Under Changing Environments. **Allergy, Asthma & Immunology Research**, v. 11, n. 4, p. 450, 2019.

AGLAS, L. et al. In vivo Induction of Functional Inhibitory IgG Antibodies by a Hypoallergenic Bet v 1 Variant. **Frontiers in Immunology**, v. 11, n. September, p. 1–15, 2020.

AHMAD, I. et al. Overcoming challenges for amplified expression of recombinant proteins using *Escherichia coli.* **Protein Expression and Purification**, v. 144, n. October 2017, p. 12–18, 2018.

AKDIS, C. A. Therapies for allergic inflammation: Refining strategies to induce tolerance. **Nature Medicine**, v. 18, n. 5, p. 736–749, 2012.

AKDIS, C. A.; AKDIS, M. Advances in allergen immunotherapy: Aiming for complete tolerance to allergens. **Science Translational Medicine**, v. 7, n. 280, p. 1–6, 2015.

AKINFENWA, O. et al. Novel vaccines for allergen-specific immunotherapy. **Current opinion in allergy and clinical immunology**, v. 21, n. 1, p. 86–99, 2021.

ALVES, A. M. et al. Clinical features and associated factors with severe asthma in salvador, Brazil. **Jornal Brasileiro de Pneumologia**, v. 46, n. 3, p. 1–8, 2020.

ANSARI, H. R.; RAGHAVA, G. P. Identification of conformational B-cell Epitopes in an antigen from its primary sequence. **Immunome Research**, v. 6, n. 1, p. 6, 2010.

ANSOTEGUI, I. J. et al. A WAO — ARIA — GA2LEN consensus document on molecular-based allergy diagnosis (PAMD@): Update 2020. **World Allergy Organization Journal**, v. 13, n. 2, p. 100091, 2020.

BENKERT, P.; TOSATTO, S. C. E.; SCHOMBURG, D. QMEAN: A comprehensive scoring function for model quality assessment. **Proteins: Structure, Function, and Bioinformatics**, v. 71, n. 1, p. 261–277, 2008.

BERENDSEN, H. J. C. et al. Molecular dynamics with coupling to an external bath. **The Journal of Chemical Physics**, v. 81, n. 8, p. 3684–3690, 1984.

BERG, L. et al. The expression of recombinant genes in *Escherichia coli* can be strongly stimulated at the transcript production level by mutating the DNA-region corresponding to the 5′-untranslated part of mRNA. **Microbial Biotechnology**, v. 2,

n. 3, p. 379–389, 2009.

BERNARDINI, R. et al. Cross-reactivity between IgE-binding proteins from Anisakis simplex and Dermatophagoides pteronyssinus. **International journal of immunopathology and pharmacology.**, v. 18, n. 4, p. 671–675, 2005.

BESSOT, J. C.; PAULI, G. Mite allergens: An overview. **European Annals of Allergy and Clinical Immunology**, v. 43, n. 5, p. 141–156, 2011.

BHANDARI, B. K.; GARDNER, P. P. TISIGNER . com : web services for improving recombinant protein production. v. 49, n. March, p. 654–661, 2021.

BHANDARI, B. K.; LIM, C. S.; GARDNER, P. P. TISIGNER.com: Web services for improving recombinant protein production. **Nucleic Acids Research**, v. 49, n. W1, p. W654–W661, 2021.

BILL, R. M. Playing catch-up with *Escherichia coli:* Using yeast to increase success rates in recombinant protein production experiments. **Frontiers in Microbiology**, v. 5, n. MAR, p. 1–5, 2014.

BLANK, S.; HILGER, C. Editorial: Novel Advances in Allergy Diagnosis and Treatment. **Frontiers in Immunology**, v. 12, n. April, p. 10–12, 2021.

BROOKS, S. M.; ALPER, H. S. Applications, challenges, and needs for employing synthetic biology beyond the lab. **Nature Communications**, v. 12, n. 1, p. 1–16, 2021.

CALDERÓN, M. A. et al. Respiratory allergy caused by house dust mites: What do we really know? **Journal of Allergy and Clinical Immunology**, v. 136, n. 1, p. 38–48, 2015.

CAMPANA, R. et al. Vaccination of nonallergic individuals with recombinant hypoallergenic fragments of birch pollen allergen Bet v 1: Safety, effects, and mechanisms. **Journal of Allergy and Clinical Immunology**, v. 143, n. 3, p. 1258–1261, 2019.

CARVALHO, K. DOS A. et al. Blomia tropicalis Blo t 5 and Blo t 21 recombinant allergens might confer higher specificity to serodiagnostic assays than whole mite extract. **BMC Immunology**, v. 14, n. 1, 2013.

CASALE, T. B.; STOKES, J. R. Immunotherapy: What lies beyond. **Journal of Allergy and Clinical Immunology**, v. 133, n. 3, p. 612–619, 2014.

CHAN, S. L. et al. Nuclear Magnetic Resonance Structure and IgE Epitopes of Blo t 5, a Major Dust Mite Allergen. **The Journal of Immunology**, v. 181, n. 4, p. 2586–2596, 2008.

CHEATHAM, T. E. I. I. I. et al. Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. **Journal of the American Chemical Society**, v. 117, n. 14, p. 4193–4194, abr. 1995.

CHEN, H. et al. In silico prediction of the T-cell and IgE-binding epitopes of Per a 6 and Bla g 6 allergens in cockroaches. **Molecular Medicine Reports**, v. 10, n. 4, p. 2130–2136, 2014.

CHEN, K. W. et al. Hypoallergenic der p 1/Der p 2 combination vaccines for immunotherapy of house dust mite allergy. **Journal of Allergy and Clinical Immunology**, v. 130, n. 2, p. 435- 443.e4, 2012.

CHEN, R. Bacterial expression systems for recombinant protein production: *E. coli* and beyond. **Biotechnology Advances**, v. 30, n. 5, p. 1102–1107, 2012.

CHIZZOLINI, F. et al. Gene position more strongly influences cell-free protein expression from operons than T7 transcriptional promoter strength. **ACS Synthetic Biology**, v. 3, n. 6, p. 363–371, 2014.

CONDIC-JURKIC, K. et al. The reliability of molecular dynamics simulations of the multidrug transporter P-glycoprotein in a membrane environment. **PLoS ONE**, v. 13, n. 1, p. 1–24, 2018.

CONSORTIUM, T. U. UniProt: a worldwide hub of protein knowledge. **Nucleic Acids Research**, v. 47, n. D1, p. D506–D515, jan. 2019.

CORREA, A.; OPPEZZO, P. Tuning different expression parameters to achieve soluble recombinant proteins in *E. coli*: Advantages of high-throughput screening. **Biotechnology Journal**, v. 6, n. 6, p. 715–730, 2011.

CRUZ, A. A. et al. The impact of a program for control of asthma in a low-income setting. **World Allergy Organization Journal**, v. 3, n. 4, p. 167–174, 2010.

CURIN, M. et al. Similar localization of conformational IgE epitopes on the house dust mite allergens Der p 5 and Der p 21 despite limited IgE cross-reactivity. **Allergy: European Journal of Allergy and Clinical Immunology**, v. 73, n. 8, p. 1653–1661, 2018.

DA SANTIAGO, H. C. et al. Structural differences between human proteins and aero- and microbial allergens define allergenicity. **PLoS ONE**, v. 7, n. 7, 2012.

DA SILVA, E. S. et al. A hybrid of two major Blomia tropicalis allergens as an allergy vaccine candidate. **Clinical and Experimental Allergy**, v. 50, n. 7, p. 835–847, 2020.

DANA, R. Advances and innovations in recombinant protein expression technology. **Journal of Advanced Pharmacy Education and Research**, v. 8, p. 88–91, 2018.

DARBY, R. A. J.; HINE, A. V. LacI-mediated sequence-specific affinity purification of plasmid DNA for therapeutic applications. **The FASEB Journal**, v. 19, n. 7, p. 1–20, 2005.

DE ARAUJO CARDOSO, T. et al. The impact of asthma in Brazil: A longitudinal analysis of data from a Brazilian national database system. **Jornal Brasileiro de Pneumologia**, v. 43, n. 3, p. 163–168, 2017.

DOROFEEVA, Y. et al. Past, present, and future of allergen immunotherapy vaccines. **Allergy: European Journal of Allergy and Clinical Immunology**, v. 76, n. 1, p. 131–149, 2021.

DRAGOSITS, M.; NICKLAS, D.; TAGKOPOULOS, I. A synthetic biology approach to self-regulatory recombinant protein production in *Escherichia coli.* **Journal of Biological Engineering**, v. 6, n. 1, p. 2, 2012.

EICHHORN, S. et al. Rational Design, Structure–Activity Relationship, and Immunogenicity of Hypoallergenic Pru p 3 Variants. **Molecular Nutrition and Food Research**, v. 63, n. 18, p. 1–10, 2019.

FENG, Y. et al. A novel self-cleavage system for production of soluble recombinant protein in *Escherichia coli.* **Protein expression and purification**, v. 99C, p. 64–69, 2014.

FERNANDES, S. DE S. C. et al. Epidemiological trends of allergic diseases in adolescents. **Jornal Brasileiro de Pneumologia**, v. 43, n. 5, p. 368–372, 2017.

FERREIRA, F.; WOLF, M.; WALLNER, M. Molecular approach to allergy diagnosis and therapy. **Yonsei Medical Journal**, v. 55, n. 4, p. 839–852, 2014.

FIRAT DUZENLI, O.; OKAY, S. Promoter engineering for the recombinant protein production in prokaryotic systems. **AIMS Bioengineering**, v. 7, n. 2, p. 62–81, 2020.

GAO, Y. F. et al. Identification and characterization of a novel allergen from Blomia tropicalis: Blo t 21. **Journal of Allergy and Clinical Immunology**, v. 120, n. 1, p. 105–112, 2007.

GARCÍA-FRUITÓS, E. Insoluble proteins: Methods and protocols. **Insoluble Proteins: Methods and Protocols**, v. 1258, p. 1–422, 2014.

GARRIDO-ARANDIA, M. et al. Molecular Dynamics of Major Allergens from Alternaria, Birch Pollen and Peach. **Molecular Informatics**, v. 33, n. 10, p. 682–694, 2014.

GILL, M.; MCCULLY, M. E. Molecular dynamics simulations suggest stabilizing mutations in a de novo designed α/β protein. **Protein Engineering, Design and Selection**, v. 32, n. 7, p. 317–329, 2019.

GUSTAFSSON, C. et al. Engineering genes for predictable protein expression. **Protein Expression and Purification**, v. 83, n. 1, p. 37–46, 2012.

GUSTAFSSON, C.; GOVINDARAJAN, S.; MINSHULL, J. Codon bias and heterologous protein expression. **Trends in Biotechnology**, v. 22, n. 7, p. 346–353, 2004.

HEBDITCH, M. et al. Protein-Sol: A web tool for predicting protein solubility from sequence. **Bioinformatics**, v. 33, n. 19, p. 3098–3100, 2017.

HOFER, H. et al. Tackling Bet v 1 and associated food allergies with a single hybrid protein. **Journal of Allergy and Clinical Immunology**, v. 140, n. 2, p. 525- 533.e10,

2017.

HON, J. et al. SoluProt: prediction of soluble protein expression in *Escherichia coli* . **Bioinformatics**, v. 37, n. 1, p. 23–28, 2021.

HUANG, J.; MACKERELL, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. **Journal of Computational Chemistry**, v. 34, n. 25, p. 2135–2145, 2013.

KAHL, L. J.; ENDY, D. A survey of enabling technologies in synthetic biology. **Journal of Biological Engineering**, v. 7, n. 1, p. 1, 2013.

KAUR, J.; KUMAR, A.; KAUR, J. Strategies for optimization of heterologous protein expression in *E. coli:* Roadblocks and reinforcements. **International Journal of Biological Macromolecules**, v. 106, p. 803–822, 2018.

KHALAF, N. F. et al. Indoor house dust-borne fungi and risk of allergic respiratory diseases in Baghdad city. **Revue Francaise d'Allergologie**, v. 62, n. 4, p. 401–406, 2022.

KHALIL, A. S.; COLLINS, J. J. Synthetic biology: Applications come of age. **Nature Reviews Genetics**, v. 11, n. 5, p. 367–379, 2010.

KI, M. R.; PACK, S. P. Fusion tags to enhance heterologous protein expression. **Applied Microbiology and Biotechnology**, v. 104, n. 6, p. 2411–2425, 2020.

KIDON, M. I. et al. Mite component-specific IgE repertoire and phenotypes of allergic disease in childhood: The tropical perspective. **Pediatric Allergy and Immunology**, v. 22, n. 2, p. 202–210, 2011.

KIM, C. R. et al. Cross-reactivity between group-5 and-21 mite allergens from Dermatophagoides farinae, Tyrophagus putrescentiae and Blomia tropicalis. **Molecular Medicine Reports**, v. 12, n. 4, p. 5467–5474, 2015.

KLEFFNER, R. et al. Foldit Standalone: a video game-derived protein structure manipulation interface using Rosetta. **Bioinformatics (Oxford, England)**, v. 33, n. 17, p. 2765–2767, 2017.

KOLASKAR, A. S.; TONGAONKAR, P. C. A semi-empirical method for prediction of antigenic determinants on protein antigens. **FEBS Letters**, v. 276, n. 1–2, p. 172–174, 1990.

KOMAR, A. A. The Art of Gene Redesign and Recombinant Protein Production: Approaches and Perspectives. In: **Topics in Medicinal Chemistry**. [s.l.] Springer Berlin Heidelberg, 2016. p. 1–17.

KONDO, T.; YUMURA, S. Strategies for enhancing gene expression in *Escherichia coli.* **Applied Microbiology and Biotechnology**, v. 104, n. 9, p. 3825–3834, 2020.

KOSURI, S. et al. Composability of regulatory sequences controlling transcription and translation in *Escherichia coli.* **Proceedings of the National Academy of Sciences of the United States of America**, v. 110, n. 34, p. 14024–14029, 2013.

KRINGELUM, J. V. et al. Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking. **PLoS Computational Biology**, v. 8, n. 12, 2012.

KWOK, R. Five hard truths for synthetic biology. **Nature**, v. 463, n. 7279, p. 288–290, 2010.

LAIMER, J. et al. MAESTRO--multi agent stability prediction upon point mutations. **BMC bioinformatics**, v. 16, n. 1, p. 116, 2015.

LI, X. et al. In silico prediction of T and B cell epitopes of der f 25 in Dermatophagoides farinae. **International Journal of Genomics**, v. 2014, 2014.

LI, Z. J. et al. CRISPR-Based Construction of a BL21(DE3)-Derived Variant Strain Library to Rapidly Improve Recombinant Protein Production. **ACS Synthetic Biology**, v. 11, n. 1, p. 343–352, 2022.

LIMÃO, R. et al. Molecular Sensitization Profile to Dermatophagoides pteronyssinus Dust Mite in Portugal. **Journal of Investigational Allergology and Clinical Immunology**, v. 32, n. 1, p. 1–18, 2020.

LOZANO TEROL, G. et al. Impact of the Expression System on Recombinant Protein Production in *Escherichia coli* BL21. **Frontiers in Microbiology**, v. 12, n. June, p. 1–12, 2021.

MACHADO, Y. et al. Fold stability during endolysosomal acidification is a key factor for allergenicity and immunogenicity of the major birch pollen allergen. **Journal of Allergy and Clinical Immunology**, v. 137, n. 5, p. 1525–1534, 2016.

MARI, A. et al. Allergen databases: Current status and perspectives. **Current Allergy and Asthma Reports**, v. 9, n. 5, p. 376–383, 2009.

MAROGNA, M. et al. Long-lasting effects of sublingual immunotherapy according to its duration: A 15-year prospective study. **Journal of Allergy and Clinical Immunology**, v. 126, n. 5, p. 969–975, 2010.

MARSCHALL, L.; SAGMEISTER, P.; HERWIG, C. Tunable recombinant protein expression in *E. coli:* enabler for continuous processing? **Applied Microbiology and Biotechnology**, v. 100, n. 13, p. 5719–5728, 2016.

MARTÍNEZ, D. et al. An engineered hybrid protein from dermatophagoides pteronyssinus allergens shows hypoallergenicity. **International Journal of Molecular Sciences**, v. 20, n. 12, 2019.

MEE, C.; BANKI, M. R.; WOOD, D. W. Towards the elimination of chromatography in protein purification: Expressing proteins engineered to purify themselves. **Chemical Engineering Journal**, v. 135, n. 1–2, p. 56–62, 2008.

MEGHJI, J. et al. Improving lung health in low-income and middle-income countries: from challenges to solutions. **The Lancet**, v. 397, n. 10277, p. 928–940, 2021.

MENZELLA, H. G. Comparison of two codon optimization strategies to enhance

recombinant protein production in *Escherichia coli.* **Microbial Cell Factories**, v. 10, p. 11–15, 2011.

MICHAEL HOLMBOE. **Import and plot Gromacs .xvg data files**.

MIKIEWICZ, D. et al. Novel Expression Vectors Based on the pIGDM1 Plasmid. **Molecular Biotechnology**, v. 61, n. 10, p. 763–773, 2019.

MILLER, J. D. The Role of Dust Mites in Allergy. **Clinical Reviews in Allergy and Immunology**, v. 57, n. 3, p. 312–329, 2019.

MIRZADEH, K. et al. Enhanced Protein Production in *Escherichia coli* by Optimization of Cloning Scars at the Vector-Coding Sequence Junction. **ACS Synthetic Biology**, v. 4, n. 9, p. 959–965, 2015.

MISTRY, J. et al. Pfam: The protein families database in 2021. **Nucleic Acids Research**, v. 49, n. D1, p. D412–D419, 2021.

MUTALIK, V. K. et al. Precise and reliable gene expression via standard transcription and translation initiation elements. **Nature Methods**, v. 10, n. 4, p. 354–360, 2013a.

MUTALIK, V. K. et al. Quantitative estimation of activity and quality for collections of functional genetic elements. **Nature Methods**, v. 10, n. 4, p. 347–353, 2013b.

NA, D.; LEE, D. RBSDesigner: Software for designing synthetic ribosome binding sites that yields a desired level of protein expression. **Bioinformatics**, v. 26, n. 20, p. 2633–2634, 2010.

NA, D.; LEE, S.; LEE, D. Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes. **BMC Systems Biology**, v. 4, 2010.

NAJAFI, N. et al. Fusion proteins consisting of Bet v 1 and Phl p 5 form IgE-reactive aggregates with reduced allergenic activity. **Scientific Reports**, v. 9, n. 1, p. 1–12, 2019.

PACHECO, B. et al. A screening strategy for heterologous protein expression in *Escherichia coli* with the highest return of investment. **Protein Expression and Purification**, v. 81, n. 1, p. 33–41, 2012.

PACKIAM, K. A. R. et al. Stepwise optimization of recombinant protein production in *Escherichia coli* utilizing computational and experimental approaches. **Applied Microbiology and Biotechnology**, v. 104, n. 8, p. 3253–3266, 2020.

PANG, S. L. et al. Crystal structure and epitope analysis of house dust mite allergen Der f 21. **Scientific Reports**, v. 9, n. 1, p. 1–13, 2019.

PAUL, S. et al. Selection of a T7 promoter mutant with enhanced in vitro activity by a novel multi-copy bead display approach for in vitro evolution. **Nucleic Acids Research**, v. 41, n. 1, p. 1–11, 2013.

PAUL, S. et al. Determination of a Predictive Cleavage Motif for Eluted Major

Histocompatibility Complex Class II Ligands. **Frontiers in immunology**, v. 9, p. 1795, 2018.

PAWANKAR, R. It's time for an evolution. **Asia Pacific Allergy**, v. 11, n. 1, p. 11–13, 2021.

PIKKEMAAT, M. G. et al. Molecular dynamics simulations as a tool for improving protein stability. **Protein Engineering**, v. 15, n. 3, p. 185–192, 2002.

PINHEIRO, C. S. et al. En route to personalized medicine: uncovering distinct IgE reactivity pattern to house dust mite components in Brazilian and Austrian allergic patients. **Clinical and Translational Allergy**, v. 11, n. 1, p. 2–5, 2021.

PONOMARENKO, J. et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. **BMC bioinformatics**, v. 9, p. 514, 2008.

POSA, D. et al. Evolution and predictive value of IgE responses toward a comprehensive panel of house dust mite allergens during the first 2 decades of life. **Journal of Allergy and Clinical Immunology**, v. 139, n. 2, p. 541- 549.e8, 2017.

PUETZ, J.; WURM, F. M. Recombinant Proteins for Industrial versus Pharmaceutical Purposes: A Review of Process and Pricing. **Processes**, v. 7, n. 8, p. 476, 2019.

PUIGBÒ, P. et al. OPTIMIZER: A web server for optimizing the codon usage of DNA sequences. **Nucleic Acids Research**, v. 35, n. SUPPL.2, p. 126–131, 2007.

PULSAWAT, P. et al. Characterization of the house dust mite allergen der p 21 produced in Pichia pastoris. **Protein Expression and Purification**, v. 101, p. 8–13, 2014.

RAPIN, N. et al. Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. **PLoS ONE**, v. 5, n. 4, 2010.

ROBERTSON, M. J.; TIRADO-RIVES, J.; JORGENSEN, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. **Journal of Chemical Theory and Computation**, v. 11, n. 7, p. 3499–3509, 2015.

RØKKE, G. et al. BioBrick assembly standards and techniques and associated software tools. **Methods in Molecular Biology**, v. 1116, p. 1–24, 2014.

ROSANO, G. L.; CECCARELLI, E. A. Recombinant protein expression in *Escherichia coli:* Advances and challenges. **Frontiers in Microbiology**, v. 5, n. APR, p. 1–17, 2014.

ROSANO, G. L.; MORALES, E. S.; CECCARELLI, E. A. New tools for recombinant protein production in *Escherichia coli*: A 5-year update. **Protein Science**, v. 28, n. 8, p. 1412–1422, 2019.

SADE, K.; ROITMAN, D.; KIVITY, S. Sensitization to dermatophagoides, Blomia tropicalis, and other mites in atopic patients. **Journal of Asthma**, v. 47, n. 8, p. 849–852, 2010.

SADEGHIAN-RIZI, T. et al. Improvement of solubility and yield of recombinant protein expression in *E. coli* using a two-step system. **Research in Pharmaceutical Sciences**, v. 14, n. 5, p. 400–407, 2019.

SAHA, S.; RAGHAVA, G. P. S. BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. **Artificial Immune Systems: Third international conference, ICARIS 2004, Catania Sicily, Italy, September 13-16, 2004 Proceedings**, p. 197–204, 2004.

SALIS, H. M. The ribosome binding site calculator. **Methods in Enzymology**, v. 498, p. 19–42, 2011.

SALIS, H. M.; MIRSKY, E. A.; VOIGT, C. A. Automated design of synthetic ribosome binding sites to control protein expression. **Nature Biotechnology**, v. 27, n. 10, p. 946–950, 2009.

SANTOS, S. P. O. et al. Engineering an optimized expression operating unit for improved recombinant protein production in *Escherichia coli*. **Protein Expression and Purification**, v. 199, n. July, 2022a.

SANTOS, S. P. O. et al. Rationally designed hypoallergenic mutant variants of the house dust mite allergen Der p 21. **Biochimica et Biophysica Acta - General Subjects**, v. 1866, n. 4, 2022b.

SCHLEGEL, S.; GENEVAUX, P.; DE GIER, J. W. Isolating *Escherichia coli* strains for recombinant protein production. **Cellular and Molecular Life Sciences**, v. 74, n. 5, p. 891–908, 2017.

SCHLESINGER, O. et al. Tuning of Recombinant Protein Expression in *Escherichia coli* by Manipulating Transcription, Translation Initiation Rates, and Incorporation of Noncanonical Amino Acids. **ACS Synthetic Biology**, v. 6, n. 6, p. 1076–1085, 2017.

SEO, S. W. et al. Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. **Metabolic Engineering**, v. 15, n. 1, p. 67–74, 2013.

SHILLING, P. J. et al. Improved designs for pET expression plasmids increase protein production yield in *Escherichia coli*. **Communications Biology**, v. 3, n. 1, 2020.

SILVA, E. S. DA et al. Advances in patent applications related to allergen immunotherapy. **Expert Opinion on Therapeutic Patents**, v. 26, n. 6, p. 657–668, 2016.

SILVA, E. S. DA et al. Dermatophagoides spp. hypoallergens design: what has been achieved so far? **Expert Opinion on Therapeutic Patents**, v. 30, n. 3, p. 163–177, 2020.

SILVA, F. S. R. et al. In vivo cleavage of solubility tags as a tool to enhance the levels of soluble recombinant proteins in *Escherichia coli*. **Biotechnology and Bioengineering**, v. 118, n. 11, p. 4159–4167, 2021.

SIRCAR, G. et al. Epitope mapping of Rhi o 1 and generation of a hypoallergenic

variant. **Journal of Biological Chemistry**, v. 291, n. 34, p. 18016–18029, 2016.

SONG, Y. et al. High-Resolution Comparative Modelingwith RosettaCM. **Structure**, v. 21, n. 10, p. 1735–1742, 2013.

TAN, S. I.; HSIANG, C. C.; NG, I. S. Tailoring Genetic Elements of the Plasmid-Driven T7 System for Stable and Robust One-Step Cloning and Protein Expression in Broad *Escherichia coli*. **ACS Synthetic Biology**, v. 10, n. 10, p. 2753–2762, 2021.

TAN, X. et al. Synthetic biology in the clinic: engineering vaccines, diagnostics, and therapeutics. **Cell**, v. 184, n. 4, p. 881–898, 2021.

THALHAMER, T. et al. Designing hypoallergenic derivatives for allergy treatment by means of in silico mutation and screening. **Journal of Allergy and Clinical Immunology**, v. 125, n. 4, 2010.

TRIPATHI, N. K.; SHRIVASTAVA, A. Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. **Frontiers in Bioengineering and Biotechnology**, v. 7, n. December, 2019.

TSCHEPPE, A.; BREITENEDER, H. Recombinant Allergens in Structural Biology, Diagnosis, and Immunotherapy. **International Archives of Allergy and Immunology**, v. 172, n. 4, p. 187–202, 2017.

TUNGEKAR, A. A.; CASTILLO-CORUJO, A.; RUDDOCK, L. W. So you want to express your protein in *Escherichia coli*? **Essays in Biochemistry**, v. 65, n. 2, p. 247–260, 2021.

VALENTA, R. et al. Recombinant allergens for allergen-specific immunotherapy: 10 years anniversary of immunotherapy with recombinant allergens. **Allergy: European Journal of Allergy and Clinical Immunology**, v. 66, n. 6, p. 775–783, 2011.

VALENTA, R. et al. Vaccine development for allergen-specific immunotherapy based on recombinant allergens and synthetic allergen peptides: Lessons from the past and novel mechanisms of action for the future. **Journal of Allergy and Clinical Immunology**, v. 137, n. 2, p. 351–357, 2016.

VALENTA, R.; CAMPANA, R.; NIEDERBERGER, V. Recombinant allergy vaccines based on allergen-derived B cell epitopes. **Immunology Letters**, v. 189, p. 19–26, 2017.

VOGL, T.; HARTNER, F. S.; GLIEDER, A. New opportunities by synthetic biology for biopharmaceutical production in Pichia pastoris. **Current Opinion in Biotechnology**, v. 24, n. 6, p. 1094–1101, 2013.

VRTALA, S. et al. Strategies for converting allergens into hypoallergenic vaccine candidates. **Methods**, v. 32, n. 3, p. 313–320, 2004.

WALLNER, M. et al. Reshaping the Bet v 1 fold modulates TH polarization. **Journal of Allergy and Clinical Immunology**, v. 127, n. 6, p. 1–19, 2011.

WALLNER, M.; PICHLER, U.; FERREIRA, F. Recombinant allergens for pollen

immunotherapy. **Immunotherapy**, v. 5, n. 12, p. 1323–1338, 2013.

WEGHOFER, M. et al. Characterization of Der p 21, a new important allergen derived from the gut of house dust mites. **Allergy: European Journal of Allergy and Clinical Immunology**, v. 63, n. 6, p. 758–767, 2008.

WELCH, M. et al. Design parameters to control synthetic gene expression in *Eschorichia coli*. **PLoS ONE**, v. 4, n. 9, 2009.

YOUNG, E.; ALPER, H. Synthetic biology: Tools to design, build, and optimize cellular processes. **Journal of Biomedicine and Biotechnology**, v. 2010, 2010.

YU, S. J.; LIAO, E. C.; TSAI, J. J. Effects of local nasal immunotherapy in allergic airway inflammation: Using urea denatured Dermatophagoides pteronyssinus. **Human Vaccines and Immunotherapeutics**, v. 11, n. 4, p. 915–921, 2015.

ZABORSKY, N. et al. Antigen Aggregation Decides the Fate of the Allergic Immune Response. **The Journal of Immunology**, v. 184, n. 2, p. 725–735, 2010.

ZHANG, D.; LAZIM, R. Application of conventional molecular dynamics simulation in evaluating the stability of apomyoglobin in urea solution. **Scientific Reports**, v. 7, n. March, p. 1–12, 2017.

ZHANG, W. et al. Computational prediction of conformational B-cell epitopes from antigen primary structures by ensemble learning. **PLoS ONE**, v. 7, n. 8, p. 1–9, 2012.

ZHANG, Y. et al. Efficient Molecular Biological Manipulations with Improved Strategies Based on Novel *Escherichia coli* Vectors. **Journal of Agricultural and Food Chemistry**, v. 69, n. 17, p. 5086–5095, 2021.

ZHERNOV, Y. et al. Recombinant allergens for immunotherapy: State of the art. **Current Opinion in Allergy and Clinical Immunology**, v. 19, n. 4, p. 402–414, 2019.

ZHOU, S. et al. Obtaining a Panel of Cascade Promoter-5′-UTR Complexes in *Escherichia coli.* **ACS Synthetic Biology**, v. 6, n. 6, p. 1065–1075, 2017.

8 APÊNDICE

8.1 PATENTE

Patente BR 10 2020 024449 3 - Submetida ao INPI
Data do Depósito: 30/11/2020
Data da Publicação Nacional: 14/06/2022

Título: Processo de construção e produção de proteínas recombinantes hipoalergênicas para uso profilático e terapêutico de doenças alérgicas causadas pelo ácaro da poeira doméstica *Dermatophagoides pteronyssinus*

INVENTORES: SARA PATRÍCIA DE OLIVEIRA SANTOS/ AYRTON BRENO PIMENTA LISBOA/ FILIPE SAMPAIO REIS DA SILVA/ LUIS FABIÁN SALAZAR GARCÉS/ LEONARDO FREIRE SANTIAGO/ EDUARDO SANTOS DA SILVA/ CARINA DA SILVA PINHEIRO/ NEUZA MARIA ALCANTARA-NEVES/ LUIS GUSTAVO CARVALHO PACHECO

RESUMO

Patente de invenção: **“Processo de construção e produção de proteínas recombinantes hipoalergênicas para uso profilático e terapêutico de doenças alérgicas causadas pelo ácaro da poeira doméstica *Dermatophagoides pteronyssinus*”.**

A presente invenção relata a construção e produção de proteínas recombinantes hipoalergênicas derivadas do alérgeno Der p 21 de *Dermatophagoides pteronyssinus,* tendo potencial uso em composições de vacinas para profilaxia e terapia de doenças alérgicas causadas por esse ácaro.

Mais particularmente a invenção refere-se à imunoterapia alérgeno-específica para as doenças alérgicas mediadas pela imunoglobulina E (IgE) e causadas pelo ácaro da poeira doméstica *Dermatophagoides pteronyssinus.*

Mais especificamente, trate-se de quatro polipeptídeos engenheirados a partir do alérgeno tipo-selvagem Der p 21, cujas sequências estão apresentadas em SEQ ID N°.: 1, SEQ ID N°.: 2, SEQ ID N°.: 3 e SEQ ID N°.: 4, caracterizados por uma redução na reatividade de IgE em soros de indivíduos alérgicos.

Também se refere aos métodos de produção, purificação e caracterização imunológica dos polipeptídeos apresentados em SEQ ID N°.: 1, SEQ ID N°.: 2, SEQ ID N°.: 3 e SEQ ID N°.: 4.

RELATÓRIO DESCRITIVO

Campo de aplicação

[1] A presente invenção refere-se à construção e produção de proteínas recombinantes hipoalergênicas derivadas da sequência do alérgeno Der p 21, presente no ácaro da poeira doméstica *Dermatophagoides pteronyssinus*.

[2] Mais particularmente a invenção relaciona-se com a imunoterapia alérgeno-específica de doenças alérgicas causadas pelo ácaro *Dermatophagoides pteronyssinus*; metodologia indutora de tolerância para as doenças alérgicas mediadas pela Imunoglobulina E (IgE).

[3] Especificamente, a invenção trata-se de quatro proteínas hipoalergênicas derivadas da sequência do Der p 21 tipo-selvagem, que foram submetidas à mutagênese *in silico* de resíduos de aminoácidos das regiões dos epítopos de células B, com o objetivo de reduzir a capacidade de ligação com a IgE.

Estado da técnica

[4] As doenças alérgicas são as enfermidades imunomediadas mais comuns em todo o mundo, as quais tiveram suas prevalências aumentadas ao longo das últimas décadas, atingindo proporções epidêmicas (BOUSQUET, J.; DAHL, R. e KHALTAEV, N. Global alliance against chronic respiratory diseases. Allergy, v. 62, n. 3, p. 216–223, 2007; CAMPBELL, D. E. e MEHR, S. Fifty years of allergy: 1965-2015. J Paediatr Child Health, v. 51, n. 1, p. 91-93, 2015). Mais de 25% da população mundial tem algum tipo de doença alérgica como asma, rinite, dermatite atópica e alergia alimentar (SCHEURER, S.; TODA, M. e VIETHS, S. What makes an allergen? Clin Exp Allergy, v. 45, n. 7, p. 1150–1161, 2015; LINHART, B. e VALENTA, R. Vaccines for allergy. Curr Opin Immunol, v. 24, n. 3, p. 354–360, 2012). De acordo com as estimativas da Organização Mundial da Saúde (OMS) em 2015, cerca de 400 milhões de pessoas no mundo são afetadas pela rinite alérgica e sua prevalência continua aumentando. A situação da asma é semelhante, uma doença que afeta atualmente mais de 300 milhões de pessoas no mundo, com previsão de

que em 2025 esse valor chegue a 400 milhões de pacientes (PAWANKAR, R. Allergic diseases and asthma: a global public health concern and a call to action. World Allergy Organ J, v. 7, n. 1, p. 12, 2014).

[5] A ocorrência dessa alta prevalência das doenças alérgicas faz com que elas sejam consideradas um problema de saúde pública que diminui a qualidade de vida, inclusive podendo ocasionar a morte dos pacientes (KING, R. M.; KNIBB, R. C. e HOURIHANE, J. O. Impact of peanut allergy on quality of life, stress and anxiety in the family. Allergy, v. 64, n. 3, p. 461-468, 2009). Nesse sentido, foi mostrado que em 2015 aproximadamente 397 mil pessoas foram a óbito devido a asma no mundo (GBD. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet, v. 392, n. 10159, p. 1789–1858, 2018), e a taxa de morte anual no Brasil é de cerca 2000 pessoas (CARDOSO, T. A. et al. Impacto da asma no Brasil: análise longitudinal de dados extraídos de um banco de dados governamental brasileiro. J Bras Pneumol, v. 43, n. 3, p. 163-168, 2017).

[6] As alergias são resultantes de reações de hipersensibilidade imunológica do tipo I, mediadas pela IgE, caracterizadas por uma sensibilidade excessiva do sistema imunológico a substâncias inócuas comuns no ambiente, chamadas alérgenos (SCHEURER, S.; TODA, M. e VIETHS, S. What makes an allergen? Clin Exp Allergy, v. 45, n. 7, p. 1150–1161, 2015). Os alérgenos estão presentes em muitas fontes como ácaros da poeira, pelo de animais, drogas e alimentos (PLATTS-MILLS, T. A. e WOODFOLK, J. A. Allergens and their role in the allergic immune response. Immunol Rev, v. 242, n. 1, p. 51-68, 2011).

[7] As respostas alérgicas envolvem o aumento da capacidade de linfócitos B em produzir imunoglobulina E (IgE) contra um alérgeno específico. O papel mediador da IgE nas manifestações alérgicas tem sido bem descrito, cujo ponto fundamental é a ligação de uma molécula alergênica a duas moléculas de IgE alérgeno-específicas (*cross-linking*), ligadas ao seu receptor de alta afinidade de IgE (FcεRI), presente na superfície das células efetoras. Tal ligação leva à liberação de mediadores inflamatórios provocando os sintomas imediatos de alergia. Sintomas tardios, por

sua vez, podem ser causados pelo aumento da presença e atividade de eosinófilos, e a proliferação de linfócitos T CD4+ responsáveis pelo perfil de produção e secreção das citocinas Th2 pró-inflamatórias (IL-4, IL-5 e IL-13) (GALLI, S. J.; TSAI, M. e PILIPONSKY, A. M. The development of allergic inflammation. Nature, v. 454, n. 7203, p. 445-454, 2008).

[8] Neste contexto, diversos fatores podem favorecer à indução de uma resposta alérgica. O tamanho molecular, estrutura tridimensional, carga, estabilidade e resistência à desnaturação são os principais fatores relacionados aos alérgenos (SCHEURER, S.; TODA, M. e VIETHS, S. What makes an allergen? Clin Exp Allergy, v. 45, n. 7, p. 1150–1161, 2015). Enquanto os fatores ligados ao indivíduo são a existência de doenças prévias ou fatores genéticos, como a atopia, a qual é uma tendência natural para produzir anticorpos IgE em resposta a pequenas doses de alérgenos e desenvolver sintomas alérgicos (JOHANSSON, S. G. O. et al. A revised nomenclature for allergy: An EAACI position statement from the EAACI nomenclature task force. Allergy, v. 56, n. 9, p. 813–824, 2001; POULSEN, L. K. e HUMMELSHOJ, L. Triggers of IgE class switching and allergy development. Ann Med, v. 39, n. 6, p. 440-456**,** 2007; SIRCAR, G. et al. Epitope mapping of Rhi o 1 and generation of a hypoallergenic variant. Journal of Biological Chemistry, v. 291, n. 34, p. 18016–18029, 2016).

[9] Dentre os diferentes estímulos ambientais, um dos principais fatores de risco para o desencadeamento de processos alérgicos é a sensibilização aos antígenos dos ácaros da poeira doméstica (CHUA, K. Y. et al. The *Blomia tropicalis* allergens. Protein Pept Lett, v. 14, p. 325–333, 2007; CARVALHO, K. A. et al. *Blomia tropicalis* Blo t 5 and Blo t 21 recombinant allergens might confer higher specificity to serodiagnostic assays than whole mite extract. BMC Immunology, v. 14, n. 11, 2013). Estimativas indicam que mais de 50% de todos os pacientes alérgicos do mundo sofrem de alergia a esses ácaros (WEGHOFER, M. et al. Characterization of Der p 21, a new important allergen derived from the gut of house dust mites. Allergy, v. 63, n. 6, p. 758–767, 2008). Sendo que entre os ácaros da poeira doméstica, *Blomia tropicalis*, *Dermatophagoides pteronyssinus e Dermatophagoides farinae* representam as mais importantes fontes de alérgenos em ambientes internos para pacientes com doenças alérgicas persistentes (SADE, K.; ROITMAN, D. e KIVITY, S.

Sensitization to Dermatophagoides, Blomia tropicalis, and other mites in atopic patients. J Asthma, v, 47, n. 8, p. 849-52, 2010). *Dermatophagoides spp.* são amplamente distribuídos globalmente, enquanto *Blomia tropicalis* é a espécie de ácaro mais encontrada em regiões tropicais e subtropicais do mundo (CARVALHO, K. A. et al. Blomia tropicalis Blo t 5 and Blo t 21 recombinant allergens might confer higher specificity to serodiagnostic assays than whole mite extract. BMC Immunology, v. 14, n.11, 2013).

[10] Até então, várias proteínas diferentes desses ácaros foram identificadas e caracterizadas (BESSOT, J. C. e PAULI, G. Mite allergens: an overview. European annals of allergy and clinical immunology, v. 43, n. 5, p. 141–56, 2011; CARNÉS, J. et al. Mite allergen extracts and clinical practice. Ann Allergy Asthma Immunol, v. 118, n. 3, p. 249–256, 2017; CHUA, K. Y. et al. The *Blomia tropicalis* allergens. Protein Pept Lett, v. 14, p. 325–333, 2007; THOMAS, W. R. et al. Pyroglyphid House Dust Mite Allergens. Protein Pept Lett, v. 14, n. 10, p. 943-953, 2007). Estudos descrevem que os alérgenos principais do *Blomia topicalis* são o Blo t 5 e Blo t 21 (CARVALHO, K. A. et al. Blomia tropicalis Blo t 5 and Blo t 21 recombinant allergens might confer higher specificity to serodiagnostic assays than whole mite extract. BMC Immunology, v. 14, n.11, 2013; SILVA, E. S. et al. Allergens of *Blomia tropicalis*: An Overview of Recombinant Molecules. Int Arch Allergy Immunol, v. 172, n. 4, p. 203-214, 2017), e para o *Dermatophagoides pteronyssinus* são os alérgenos Der p 1, Der p 2 e Der p 23 (CALDERON, M. A. et al. Respiratory allergy caused by house dust mites: What do we really know? J Allergy Clin Immunol, v. 136, n. 1, p. 38-48, 2015; WEGHOFER, M. et al. Identification of Der p 23, a Peritrophin-like Protein, as a New Major *Dermatophagoides pteronyssinus* Allergen Associated with the Peritrophic Matrix of Mite Fecal Pellets. J Immunol, v. 190, n. 7, p. 3059-3067, 2013).

[11] O alérgeno Der p 21 também teve sua importância clínica investigada. Ele foi caracterizado como um novo e importante alérgeno presente no intestino e nas fezes do ácaro *Dermatophagoides pteronyssinus*, apresentando alta capacidade de ligação aos anticorpos IgE e alta atividade alergênica (WEGHOFER, M. et al. Characterization of Der p 21, a new important allergen derived from the gut of house dust mites. Allergy, v. 63, n. 6, p. 758–767, 2008). Além disso, já foi descrito que a sua sequência possui semelhanças com as dos alérgenos do grupo 5, mas sem

apresentar reatividade cruzada. Alinhamentos de sequências mostraram que Der p 21 e Blo t 5 tem uma identidade de sequência de aminoácidos (AA) de 40% (BESSOT, J. C. e PAULI, G. Mite allergens: an overview. Eur Ann Allergy Clin Immunol, v. 43, n. 5, p. 141–56, 2011). Enquanto, Der p 21 e Der p 5 possuem 31% de identidade (GAO, Y. F. et al. Identification and characterization of a novel allergen from *Blomia tropicalis:* Blo t 21. J Allergy Clin Immunol, v. 120, n 1, p. 105-112, 2007), com epítopos de IgE localizados em regiões semelhantes em suas estruturas (CURIN, M. et al. Similar localization of conformational IgE epitopes on the house dust mite allergens Der p 5 and Der p 21 despite limited IgE cross-reactivity. Allergy, v. 73, n. 8, p. 1653–1661, 2018). Estudos de sequência também revelaram que Der p 21 e Blo t 21 compartilham 41% de identidade de sequência de AA (GAO, Y. F. et al. Identification and characterization of a novel allergen from Blomia tropicalis: Blo t 21. J Allergy Clin Immunol. v. 120, n 1, p. 105-112, 2007). Dessa forma, o alérgeno Der p 21 é considerado uma molécula bem definida, sendo, portanto, um forte candidato para a imunoterapia alérgeno-específica contra alergia ao ácaro *Dermatophagoides pteronyssinus*. Assim, mais estudos são necessários para a avaliação da reatividade IgE ao alérgeno Der p 21, envolvendo a produção de derivados hipoalergênicos - os quais apresentam baixa ligação de IgE e os epítopos de células T conservados.

Problemas do estado da técnica

[12] Até o momento, não existe um tratamento 100% eficaz para as doenças alérgicas. Os medicamentos disponíveis atualmente são utilizados apenas no controle dos sintomas da doença, mas não atuam diretamente na causa da doença. Além disso, as drogas antialérgicas, como os glicocorticoides e anti-histamínicos podem se tornar ineficazes e causar vários efeitos adversos (SICHERER, S. H. e LEUNG, D. Y. Advances in allergic skin disease, anaphylaxis, and hypersensitivity reactions to foods, drugs, and insects in 2012. J Allergy Clin Immunol, v. 131, n. 1, p. 55-66, 2013).

[13] Vale ressaltar, que os gastos na saúde pública para o tratamento de doenças alérgicas são grandes e crescentes, levando em conta a alta prevalência dessas doenças em escala mundial. Segundo a base de dados do Departamento de Informática do Sistema Único de Saúde (DATASUS) do Ministério da Saúde, no

Brasil há mais de 120.000 hospitalizações devido a asma anualmente, sendo a média de custo de uma hospitalização aproximadamente 160 dóláres. Durante o período de 2008 a 2013 cerca de 170 milhões de dólares foram gastos com internações hospitalares por asma (CARDOSO, T. A. et al. Impacto da asma no Brasil: análise longitudinal de dados extraídos de um banco de dados governamental brasileiro. J Bras Pneumol, v. 43, n. 3, p. 163-168, 2017).

[14] Considerando as limitações do tratamento convencional das doenças alérgicas do tipo farmacológico e os altos gastos relacionados com esse tipo de tratamento, uma abordagem terapêutica alternativa ao uso de drogas para prevenir ou tratar as doenças alérgicas é a imunoterapia alérgeno-específica (IT), utilizada há mais de 100 anos como uma metodologia terapêutica indutora de tolerância para as doenças alérgicas, agindo diretamente na causa da doença (AKDIS, C. A. e AKDIS, M. Advances in allergen immunotherapy: Aiming for complete tolerance to allergens. Sci Transl Med*,* v. 7, n. 280, p. 280–286, 2015; CASALE, T. B. e STOKES, J. R. Immunotherapy: What lies beyond. J Allergy Clin Immunol, v. 133, n. 3, p. 612–619, 2014). Esse tipo de terapia é baseada na administração de doses crescentes do alérgeno, com o objetivo de estimular tolerância ao alérgeno e diminuir a sensibilidade específica contra o mesmo (AKDIS, C. A. e AKDIS, M. Advances in allergen immunotherapy: Aiming for complete tolerance to allergens. Sci Transl Med*,* v. 7, n. 280, p. 280–286, 2015). Consequentemente, ocorre a melhora dos sintomas associados às alergias e promove uma proteção aos indivíduos contra o desenvolvimento de sintomas alérgicos durante 7 a 12 anos (CASALE, T. B. e STOKES, J. R. Immunotherapy: What lies beyond. J Allergy Clin Immunol, v. 133, n. 3, p. 612–619, 2014; MAROGNA, M. et al. Long-lasting effects of sublingual immunotherapy according to its duration: A 15-year prospective study. J Allergy Clin Immunol, v. 126, n. 5, p. 969–975, 2010).

[15] Os mecanismos de ação da imunoterapia implicam numa modulação da resposta imune do paciente, modificando o perfil predominante Th2 para uma resposta do tipo T reguladora com indução das citocinas IL-10 e TGF-β. A IL-10 induz uma mudança de isotipo de IgE para IgG4, que atua como anticorpo bloqueador da IgE, enquanto TGF-β aumenta os níveis de IgA. Um tratamento imunoterápico de sucesso é caracterizado pela redução significativa das reações

nos testes cutâneos de alergia, diminuição dos níveis de IgE e produção aumentada de IgG4 alérgeno-específica (CASALE, T. B. e STOKES, J. R. Future forms of immunotherapy. J Allergy Clin Immunol, v. 127, n. 1, p. 8-15, 2011).

Vantagens da invenção

[16] Atualmente a IT é o único tratamento curativo disponível para as doenças alérgicas; no entanto, a eficácia do tratamento depende muito da qualidade do produto alergênico aplicado (WALLNER, M.; PICHLER, U. e FERREIRA, F. Recombinant allergens for pollen immunotherapy. Immunotherapy, v. 5, n. 12, p. 1323–1338, 2013). A abordagem inicial da imunoterapia, que ainda vem sendo utilizada atualmente, baseia-se na utilização de extratos brutos, os quais consistem em misturas heterogêneas de componentes alergênicos e não-alergênicos (VRTALA, S. et al. Strategies for converting allergens into hypoallergenic vaccine candidates. Methods, v. 32, n. 3, p. 313–320, 2004; WALLNER, M.; PICHLER, U. e FERREIRA, F. Recombinant allergens for pollen immunotherapy. Immunotherapy, v. 5, n. 12, p. 1323–1338, 2013; YU, S. J.; LIAO, E. C. e TSAI, J. J. Effects of local nasal immunotherapy in allergic airway inflammation: Using urea denatured *Dermatophagoides pteronyssinus*. Hum Vaccin Immunother, v. 11, n. 4, p. 915–921, 2015). Porém, essa abordagem apresenta várias desvantagens que inviabiliza o uso de extratos para a imunoterapia em alergia, dentre elas estão: a composição molecular dos extratos pode variar consideravelmente entre as várias fontes e lotes; dificuldade na padronização do extrato; não é possível realizar um tratamento específico quando são utilizados extratos alergênicos; indução de novas sensibilizações à alérgenos, para os quais os pacientes não eram alérgicos; ausência de importantes alérgenos; eles precisam ser administrados durante longos períodos para alcançar resultados eficazes; podem causar efeitos secundários graves em pacientes, incluindo o choque anafilático que pode levar à morte (FERREIRA, F. et al. Genetic engineering of allergens for immunotherapy. Int Arch Allergy Immunol, v. 128, p. 171-178, 2002; FERREIRA, F; WOLF, M. e WALLNER, M. Molecular approach to allergy diagnosis and therapy. Yonsei Med J, v. 55, n.4, p. 839–852, 2014; MARI, A. e SCALA, E. Allergenic Extracts for Specific Immunotherapy: To Mix or Not to Mix? Int Arch Allergy Immunol, v. 141, n. 1, p. 57–

60, 2006; VRTALA, S. et al. Strategies for converting allergens into hypoallergenic vaccine candidates. Methods, v. 32, n. 3, p. 313–320, 2004).

[17] Dessa forma, esforços foram voltados para a produção de alérgenos melhor caracterizados, graças a tecnologia do DNA recombinante. Desde a clonagem dos primeiros alérgenos em 1988, a produção de moléculas recombinantes alergênicas tornou-se um procedimento de rotina, permitindo que a imunoterapia fosse realizada de forma personalizada, de acordo com o perfil de sensibilização de cada paciente (WALLNER, M.; PICHLER, U. e FERREIRA, F. Recombinant allergens for pollen immunotherapy. Immunotherapy, v. 5, n. 12, p. 1323–1338, 2013). Porém, essa alternativa resolve apenas os problemas relacionados à falta de alérgenos bem definidos dos extratos brutos, mas as moléculas recombinantes do tipo selvagem ainda apresentam a capacidade de realizar reação cruzada com a IgE pré-existente e, portanto, também podem provocar efeitos anafiláticos graves e com risco de vida (VRTALA, S. et al. Strategies for converting allergens into hypoallergenic vaccine candidates. Methods, v. 32, n. 3, p. 313–320, 2004). Para evitar esses efeitos indesejáveis e aumentar principalmente a segurança, ferramentas modernas de engenharia de proteínas têm sido utilizadas em conjunto com novas ferramentas da bioinformática com o objetivo de desenvolver moléculas recombinantes hipoalergênicas (VRTALA, S. et al. Strategies for converting allergens into hypoallergenic vaccine candidates. Methods, v. 32, n. 3, p. 313–320, 2004; THALHAMER, T. et al. Designing hypoallergenic derivatives for allergy treatment by means of in silico mutation and screening. J Allergy Clin Immunol, v. 125, n. 4, p. 926–934, 2010). Vários estudos clínicos mostram que a utilização de moléculas hipoalergênicas reduz os riscos de anafilaxia, e outros efeitos secundários, sem diminuir a eficácia quando comparada ao alérgeno recombinante não-hipoalergenizado (LINHART, B. et al. Molecular Evolution of Hypoallergenic Hybrid Proteins for Vaccination against Grass Pollen Allergy. J Immunol, v. 194, n. 8, p. 4008-4018, 2015). Isso pode ser explicado pelo fato de que as moléculas recombinantes hipoalergênicas são projetadas para que ocorra uma redução da atividade alergênica do antígeno original, por meio de mutações em epítopos de ligação à IgE, mas mantendo seus epítopos de células T. Com isso, ocorre indução de anticorpos IgG bloqueadores e não provocam efeitos colaterais graves, induzindo resposta imunológica de tolerância (SILVA, E. S. et al. Allergens of *Blomia tropicalis*:

An Overview of Recombinant Molecules. Int Arch Allergy Immunol, v. 172, n. 4, p. 203-214, 2017; THALHAMER, T. et al. Designing hypoallergenic derivatives for allergy treatment by means of in silico mutation and screening. J Allergy Clin Immunol, v. 125, n. 4, p. 926–934, 2010). Logo, alérgenos recombinantes hipoalergênicos são considerados mais adequados e seguros, o que torna possível considerar o uso de vacinas de alérgenos para prevenção e tratamento da alergia (VALENTA, R. et al. Recombinant allergens for allergen-specific immunotherapy: 10 years anniversary of immunotherapy with recombinant allergens. Allergy, v. 66, n. 6, p. 775–783, 2011).

[18] A hipoalergenização pode ser feita na base da tentativa e erro, por exemplo utilizando a metodologia do embaralhamento de DNA; realizando testes de diferentes combinações de peptídeos a partir dos respectivos antígenos recombinantes ou por métodos racionais; por meio de abordagens de mutagênese *in silico* (CHEN, K. W. et al. Hypoallergenic Der p 1/Der p 2 combination vaccines for immunotherapy of house dust mite allergy. J Allergy Clin Immunol, v. 130, n. 2, p. 435-443, 2012; THALHAMER, T. et al. Designing hypoallergenic derivatives for allergy treatment by means of in silico mutation and screening. J Allergy Clin Immunol, v. 125, n. 4, p. 926–934, 2010). Com essas novas metodologias é possível modificar a estrutura completa do alérgeno natural, bem como as regiões de epítopos reconhecidos por células B ou T, com a finalidade de reduzir sua alergenicidade e aumentar a imunogenicidade (GERALDINI, M. et al. Alérgenos recombinantes na prática da imunoterapia. Rev Bras Alergia Imunopatol, v. 31, n. 3, p. 92-97, 2008; SIRCAR, G. et al. Epitope mapping of Rhi o 1 and generation of a hypoallergenic variant. Journal of Biological Chemistry, v. 291, n. 34, p. 18016–18029, 2016; THALHAMER, T. et al. Designing hypoallergenic derivatives for allergy treatment by means of in silico mutation and screening. J Allergy Clin Immunol, v. 125, n. 4, p. 926–934, 2010). Cabe ressaltar que o uso do embaralhamento de DNA para geração de proteínas hipoalergênicas estáveis, é um processo caro e demorado, quando comparado com os métodos computacionais de reconhecimento de epítopos, que permitem a realização de mutações direcionadas e uma seleção prévia *in silico* das moléculas que serão testadas (LAIMER, J. et al. MAESTRO--multi agent stability prediction upon point mutations. BMC Bioinformatics, v. 16, n. 1, p. 116, 2015; SIRCAR, G. et al. Epitope mapping of Rhi o 1 and generation of a

hypoallergenic variant. Journal of Biological Chemistry, v. 291, n. 34, p. 18016–18029, 2016; THALHAMER, T. et al. Designing hypoallergenic derivatives for allergy treatment by means of in silico mutation and screening. J Allergy Clin Immunol, v. 125, n. 4, p. 926–934, 2010; ZHANG, W. et al. Computational prediction of conformational B-cell epitopes from antigen primary structures by ensemble learning. PLOS ONE, v. 7, n. 8, p. 1–9, 2012).

[19] Neste cenário, um estudo de Silva e colaboradores (SILVA, E. S. et al. Advances in patent applications related to allergen immunotherapy. Expert Opin Ther Pat, v. 26, n. 6, p. 657-668, 2016) revelou que as proteínas recombinantes hipoalergênicas ocupam a quarta posição no número total de pedidos de patentes nas últimas décadas, mostrando o avanço do uso de tecnologias mais direcionadas, seguras e eficazes para a imunoterapia alérgeno-específica. E ao analisar os pedidos de patentes para moléculas hipoalergênicas, observa-se que a produção de proteínas hipoalergênicas de forma racional tem se destacado como uma abordagem inovadora, apresentando um número de etapas experimentais reduzidas, o que otimiza o processo de obtenção do produto.

[20] Em outro estudo feito por Silva e colaboradores (SILVA, E. S. et al. *Dermatophagoides* spp. hypoallergens design: what has been achieved so far? Expert Opin Ther Pat, v. 30, n. 3, p. 163-177, 2020) foi feita uma busca de patentes e artigos publicados relacionados com a criação de moléculas hipoalergênicas a partir do ácaro *Dermatophagoides* spp. para produção de vacinas. Com essa busca foi verificado que os seguintes alérgenos já foram alvos para construção de vacinas hipoalergênicas: Der p 1, De r p 2, Der p 3, Der p 5, Der p 7, Der p 8, Der p 23, Der f 1, Der f 2, Der f 3, Der f 7 e Der f 13. Evidenciando que o alérgeno Der p 21 ainda não havia sido incluído em nenhum estudo de hipoalergenização, mesmo apresentando alta capacidade de ligação aos anticorpos IgE e alta atividade alergênica, como já mencionado no texto acima. Dessa forma, fica evidente a importância de um estudo que explora o alérgeno Der p 21 como alvo para produção de uma vacina hipoalergênica, como é proposto neste pedido de patente.

[21] A patente de número WO2012072678A1, de título “Hypoallergenic polypeptides for the treatment of house dust mite allergy” mostra o processo de

produção de uma proteína quimera hipoalergênica para o tratamento da alergia causada pelo ácaro *Dermatophagoides pteronyssinus*. A quimera foi produzida pela fusão das proteínas Der p 1 e Der p 2 e algumas modificações pontuais por meio do embaralhamento de DNA e mutagênese sítio dirigida. Com relação aos resultados, a quimera apresentou reatividade de IgE reduzida e baixa capacidade de ativação de basófilos isolados de pacientes alérgicos. Vale ressaltar que apesar da grande vantagem do embaralhamento de DNA de montar moléculas com potencial anafilático reduzido e imunogenicidade melhorada, essa técnica exige a necessidade de realizar vários ensaios, para finalmente definir quais são as moléculas mais adequadas, tornando assim o procedimento mais demorado. No caso do método de hipoalergenização proposto no nosso pedido de patente, as moléculas são selecionadas previamente *in silico*, antes de qualquer experimento de bancada. Apenas as moléculas com mutações mais desestabilizadoras são escolhidas para expressão heteróloga, reduzindo assim o número de etapas experimentais para a obtenção das moléculas hipoalergenizadas.

[22] A patente de número WO2004065414A1 e título “Process for the preparation of hypoallergenic mosaic antigens” mostra o processo de hipoalergenização por avaliação dos epítopos de IgE. As moléculas criadas são derivadas do alérgeno Phl p 2, do pólen da grama Timothy, e mostraram uma redução de ligação à IgE. Porém, a forma como feita realizada a reorganização da sequência dos aminoácidos impediu a manutenção da estrutura do alérgeno tipo-selvagem. No método de hipoalergenização deste pedido de patente existe uma preocupação em manter a estrutura original do alérgeno tipo-selvagem, para isso foram realizadas análises estruturais por meio de programas de bioinformática.

[23] A patente número WO2011151449 e título “Pharmaceutical product comprising mite allergen extract (s) and a method for the manufacture thereof” descreve a criação de um produto contendo o corpo inteiro e extratos fecais dos alérgenos Der p 1 e Der p 2, para o tratamento e/ou prevenção de alergias e asma alérgica causada por ácaros da poeira doméstica. Vale ressaltar que, o extrato total dos ácaros compreende mais de 30 proteínas e uma quantidade substancial de materiais não proteicos, originários do corpo ou das fezes do ácaro. Tornando seu uso

desvantajoso para a imunoterapia alérgeno-específica. O que não é o caso da nossa tecnologia, a qual se baseia apenas no uso de proteínas alergênicas.

[24] A patente número WO2016193126A1 e título “Allergy-specific immunotherapy compositions for use in the treatment of house-dust mite allergy” refere-se a uma composição contendo alérgenos Der p 1 e Der p 2 naturais purificados ou alérgenos Der f 1 e Der f 2 naturais purificados para uso no tratamento de alergias aos ácaros *D. pteronyssinus* e *D. farinae.* No entanto, a imunoterapia com alérgenos naturais podem provocar efeitos secundários, sendo mais recomendado o uso de alérgenos hipoalergênicos, os quais podem ser desenvolvidos de forma racional utilizando ferramentas de bioinformática, como aqui descrito neste pedido de patente.

[25] A patente número US8652485B2 e título “Peptide for vaccine” mostra composições apresentando combinações de fragmentos de peptídeos derivados de alérgenos de ácaros do grupo 1 (Der p 1 e Der f 1), do grupo 2 (Der p 2 e Der f 2) e do grupo 3 (Der p 7 e Der f 7), para serem utilizadas na prevenção e tratamento da alergia aos ácaros *D. pteronyssinus* e *D. farinae*. Em contrapartida, neste pedido de patente temos uma composição de um alérgeno único, o que será mais interessante quando tratarmos de casos onde a sensibilização do indivíduo for apenas para o alérgeno Der p 21.

[26] Considerando todos os dados relatados, este pedido de patente baseia-se na criação de moléculas recombinantes hipoalergênicas para a terapia de doenças alérgicas causadas pelo alérgeno Der p 21, as quais apresentam as vantagens já mencionadas acima. Vale ressaltar que o método de hipoalergenização utilizado nesta presente invenção consiste no uso de um *pipeline* de análises de bioinformática estrutural, fácil de ser executado e envolve uma abordagem de mutagênese *in silico* que permite a criação de moléculas hipoalergênicas de forma racional, sendo baseado no conhecimento prévio da estrutura tridimensional do alérgeno e da localização dos epítopos de células B, possibilitando uma modificação direcionada dos epítopos reconhecidos por IgE e mantendo suas capacidades imunogênicas. Além disso, este método aqui proposto permite uma análise das moléculas a nível computacional, o que gera resultados de quais moléculas criadas apresentam as características desejadas e devem ser utilizadas para os testes

experimentais, fazendo com que as etapas de produção sejam reduzidas. Além disso o mesmo poderá ser aplicado para a criação de outras proteínas hipoalergênicas para a imunoterapia alérgeno-específica.

Breve descrição das Figuras

[27] **Figura 1.** Estrutura tridimensional do alérgeno Der p 21 antes e depois da mutagênese *in silico.* Estrutura mostrada no modo *surface* – superfície. Com essa análise foi verificado que as mutações realizadas não alteraram a estrutura original da molécula.

[28] **Figura 2.** Resumo das construções dos plasmídeos utilizados nos experimentos da presente invenção. 1 - Clonagem dos genes sintéticos (Der p 21 wt e mutantes) no vetor pCR-Blunt II-TOPO. 2 – Sub-clonagem dos genes sintéticos (Der p 21 wt e mutantes) no vetor pOPT 2.0.

[29] **Figura 3.** Expressão do Der p 21 wt e seus mutantes na linhagem BL21(DE3), analisada em SDS-PAGE 12%. PM: Padrão de peso molecular (Biorad); 1: Der p 21 wt T20h; 2: Der p 21 wt purificado; 3: Mutante 1 T20h; 4: Mutante 1 purificado; 5: Mutante 2 T20h; 6: Mutante 2 purificado; 7: Mutante 4 T20h; 8: Mutante 4 purificado.

[30] **Figura 4.** *Western Blot* da expressão do Der p 21 wt e seus mutantes na linhagem BL21(DE3). PM: Padrão de peso molecular (Biorad); 1: Der p 21 wt T20h; 2: Mutante 1 T20h; 3: Mutante 2 T20h; 4: Mutante 4 T20h.

[31] **Figura 5.** Teste de solubilidade analisado em SDS-PAGE 12%. **(A)** Teste de solubilidade do Der p 21 wt. PM: Padrão de peso molecular (Biorad); 1: Fração Solúvel 1; 2: Fração Solúvel 2; 3: Fração Insolúvel. **(B)** Teste de solubilidade do Mutante 1 na linhagem BL21(DE3). PM: Padrão de peso molecular (Biorad); 1: Fração Solúvel 1; 2: Fração Solúvel 2; 3: Fração Insolúvel. **(C)** Teste de solubilidade do Mutante 2 na linhagem BL21(DE3). PM: Padrão de peso molecular (Biorad); 1: Fração Solúvel 1; 2: Fração Solúvel 2; 3: Fração Insolúvel. **(D)** Teste de solubilidade do mutante 4. PM: Padrão de peso molecular (Biorad); 1: Fração Solúvel 1; 2: Fração Solúvel 2; 3: Fração Insolúvel.

[32] **Figura 6.** Avaliação da reatividade de IgE por Dot blot. Um *pool* de soros de pacientes alérgicos a *Dermatophagoides pteronyssinus* (N=5) foi utilizado no ensaio. A análise da reatividade de IgE ao Der p 21 wt *versus* mutantes foi realizada utilizando o programa ImageJ. As áreas dos spots foram comparadas a nível de porcentagem, sendo que a área do spot do Der p 21 wt = 12699.075 equivale a 100% e as porcentagens dos mutantes foram calculadas em comparação com essa área (Mutante 1 área 5145.033/porcentagem 40,51%, mutante 2 área 7238.397/porcentagem 56,99%, mutante 4 área 11435.054/porcentagem 90,046%). Para o mutante 1 foi detectada uma diminuição de aproximadamente 60% na ligação de IgE nos soros de pacientes alérgicos em relação ao Der p 21 tipo-selvagem. Enquanto no ensaio com o mutante 4, foi detectada uma diminuição de apenas 10% da reatividade.

[33] **Figura 7.** Avaliação da Reatividade de IgE por ELISA indireto. Um *pool* de soros de pacientes alérgicos a *Dermatophagoides pteronyssinus* (N=5) foi utilizado no ensaio. Foi detectada uma maior reatividade de IgE para a molécula recombinante nativa em comparação com as proteínas mutantes. Dentre os mutantes, o mutante 1 que apresentou uma reatividade de IgE menor quando comparado ao Der p 21 wt, aproximadamente 67% de redução.

Descrição detalhada da tecnologia

Construção in silico dos derivados hipoalergênicos

[34] A presente invenção refere-se à construção *in silico* de moléculas hipoalergênicas a partir do alérgeno Der p 21 tipo-selvagem, por meio de substituições pontuais *in silico* de seus resíduos de aminoácidos, a fim de produzir moléculas com baixa ligação à IgE. O processo de hipoalergenização do alérgeno Der p 21 compreende os passos a seguir:

[35] **Busca por sequência e análise de domínios:** Foram feitas buscas por entradas contendo informações sobre o alérgeno Der p 21 de *Dermatophagoides pteronyssinus* no banco de dados relacional Allergome, uma plataforma para o conhecimento de alérgenos (http://www.allergome.org/). As entradas no Allergome

contêm links para as sequências no banco de dados de proteína UniProtKB (http://www.uniprot.org/) que é integrado ao banco de dados InterPro (https://www.ebi.ac.uk/interpro/), onde são encontradas informações sobre função, motivos, estruturas secundárias e peptídeo sinal.

[36] **Análise de sítios de modificações pós-traducionais:** Estudos comprovam a potencialização da alergenicidade de algumas proteínas de acordo com algumas modificações pós-traducionais. Sítios de modificações pós-traducionais foram mapeados com o servidor ScanProsite com os parâmetros padrão da plataforma (http://prosite.expasy.org/scanprosite/).

[37] **Predição da estrutura terciária:** A sequências do alérgeno Der p 21 selvagem foi submetida no servidor de predição de estrutura terciária proteica Robetta com os parâmetros padrão (modelagem *de novo*) (http://robetta.bakerlab.org/). O resultado é um arquivo de estrutura proteica em alta resolução no formato .PDB, ainda a ser refinado.

[38] **Refinamento da estrutura terciária:** O refinamento foi realizado com a plataforma interativa FoldIt *Standalone*, que refina as cadeias laterais e o esqueleto proteico baseado na energia livre/pontuação em kcal/mol. Valores mais baixos de energia livre mostram proteínas mais refinadas e estáveis. Isso significa que a proteína está mais próxima do seu estado natural.

[39] **Validação da estrutura terciária:** A avaliação e a validação da qualidade das estruturas foram realizadas no servidor QMEAN (https://swissmodel.expasy.org/qmean/). O QMEAN busca proteínas com tamanho e peso molecular semelhantes ao modelo proteico fornecido no banco de dados de proteínas RCSB (https://www.rcsb.org/pdb/home/home.do), fazendo uma comparação entre os modelos. Assim, o QMEAN estima a qualidade do modelo fornecido. Uma distribuição normal é gerada com os dados das proteínas conhecidas e o ponto de corte do servidor (valor QMEAN4) para os modelos a serem considerados validados é entre 2 e -2.

[40] **Visualização da estrutura terciária**: A visualização das estruturas 3D foi feita com o FoldIt *Standalone* e com o PyMOL *Molecular Graphics System* (https://pymol.org/).

[41] **Predição dos epítopos de células B:** Epítopos reconhecidos por célula B foram preditos nos servidores CBTope (http://osddlinux.osdd.net/raghava/cbtope/submit.php), ElliPro (http://tools.iedb.org/ellipro/), BCEPRED (http://crdd.osdd.net/raghava/bcepred/), DiscoTope (http://www.cbs.dtu.dk/services/DiscoTope/) e ElliPro (http://tools.iedb.org/ellipro/). Todos os programas foram utilizados com os parâmetros padrões. Cada servidor possui algoritmos diferentes para a predição dos epítopos, sendo assim, um consenso foi construído manualmente a partir dos resultados. Houve também uma triagem com um cut-off em cada programa. Nos servidores CBTope, BCEPRED, e DiscoTope o cut-off usado foi o do programa. No ElliPro, regiões preditas como epítopo e com escore acima de 0,6 foram selecionadas para a formação do consenso.

[42] **Mutagênese *in silico* de aminoácidos presentes nos epítopos de células B:** Após a construção do consenso, aminoácidos com escores mais altos nas listas de resultados dos servidores foram selecionados para sofrerem mutações in silico. Sendo selecionadas as posições 77 – ácido glutâmico, 82 – ácido aspártico, 87 – ácido glutâmico, 110 – lisina. Após esta seleção, foi realizado um escaneamento por mutações desestabilizantes nestas posições utilizando o servidor MAESTROweb (https://biwww.che.sbg.ac.at/maestro/web/). Por fim foram realizadas combinações entre as mutações mais desestabilizantes (E77G, D82P, E87S e K110G) para a produção das variantes do Der p 21 *wt:* Mutante 1 D82P, Mutante 2 D82P e K110G, Mutante 3 E77G, D82P e K110G, Mutante 4 E77G, D82P, E87S e K110G. Gerando as sequências das reivindicações 001: SEQ ID N°.: 1, SEQ ID N°.: 2, SEQ ID N°.: 3 e SEQ ID N°.: 4.

[43] Vale ressaltar que uma filtragem bioquímica foi feita para que não ocorresse uma possível mutagênese em locais que causassem uma ruptura completa na estrutura secundária ou, por consequência, na estrutura terciária da proteína original, como, por exemplo, uma prolina em uma alfa-hélice.

[44] **Análise das mutações na molécula tridimensional:** Uma vez selecionados os aminoácidos que substituirão os aminoácidos originais, foi feita uma mutagênese *in silico* no arquivo em .PDB gerado pelo Robetta no programa FoldIt *Standalone*, com a finalidade de observar as mudanças na estrutura da proteína (Figura 1). O resultado da mutagênese *in silico* demostrou que as variantes geradas mantiveram sua estrutura, dessa forma suas sequências foram quimicamente sintetizadas, apresentando sítios da enzima de restrição SapI nas suas extremidades.

Exemplos de utilização

[45] Para que esta invenção possa ser melhor compreendida, os exemplos seguintes são apresentados. Os exemplos são apenas para fins de ilustração e não devem ser interpretados como limitações das reivindicações feitas nesta invenção. Vale ressaltar que outros tipos de sistemas de expressão heterólogos estão disponíveis e podem ser usados para a produção das proteínas hipoalergênicas reivindicadas na presente invenção, por exemplo sistemas eucarióticos e sistemas *cell-free*. Além disso, a purificação dessas proteínas pode ser realizada por meio de várias modalidades de cromatografia e o modo de obtê-las também pode ser alterado.

Exemplo 1: Desenho e construção dos plasmídeos

[46] Os genes sintéticos codificadores do Der p 21 wt e dos mutantes foram sintetizados quimicamente. Esses genes foram primeiramente clonados no vetor pCR-Blunt II-TOPO do Kit Zero Blunt™ TOPO™ PCR Cloning (Invitrogen, Carlsbad, CA, EUA), seguindo as orientações do fabricante. Os plasmídeos foram transformados em bactérias *E. coli* XL1Blue quimiocompetentes por choque térmico, essas cepas foram então cultivadas *overnight* a 37°C em Agar Luria-Bertani (LB) - 50 µg/mL de Kanamicina e as colônias foram isoladas e inoculadas em meio LB - 50 µg/mL de Kanamicina *overnight* a 37°C. A extração dos plasmídeos foi realizada com o NucleoSpin plasmid mini-prep Kit (Macherey Nagel Co., Düren, Germany), de acordo com o protocolo do fabricante. Para confirmação das clonagens, os plasmídeos foram digeridos com a enzima de restrição EcoRI (Promega, Madison, WI, USA), de acordo com as instruções do fabricante. Os resultados das digestões

foram analisados por eletroforese em gel de agarose 1%, em tampão TAE.

[47] Os genes sintéticos foram depois sub-clonados no vetor pOPT 2.0. Para isso foi realizada uma digestão com a enzima SapI (Invitrogen, Carlsbad, CA, EUA) e os segmentos de DNA foram isolados por eletroforese em gel de agarose 1% em tampão TAE, purificados usando o *kit* de extração de DNA em gel de agarose Illustra GFX PCR DNA and Gel Band Purification (GE Healthcare, Waukesha, WI, USA) e ligados utilizando a enzima T4 DNA ligase (Promega, Madison, WI, USA). A bactéria XL1Blue também foi utilizada nesse experimento de clonagem, cultivadas com cloranfenicol 34 µg/mL. As clonagens foram confirmadas por digestão com as enzimas EcoRI e PstI (Promega, Madison, WI, USA). As construções dos plasmídeos estão resumidas na Figura 2.

Exemplo 2: Expressão em pequena escala e avaliação da solubilidade do alérgeno Der p 21 tipo-selvagem e seus derivados hipoalergênicos

[48] O vetor de expressão pOPT 2.0, contendo os genes do alérgeno Der p 21 wt e seus mutantes, foi transformado na linhagem BL21(DE3) (Invitrogen, Carlsbad, CA, EUA). Após incubação em placas LB-agar (34 µg/mL de cloranfenicol), uma única colônia resistente foi selecionada para expressão proteica, que foi realizada em 10 mL de meio LB (34 µg/mL de cloranfenicol) e a indução da expressão através da adição de 1,0 mM de Isopropil-β-D-tiogalactopiranósido (IPTG) (Invitrogen, Carlsbad, CA, EUA) quando a DO 600 nm alcançou 0,6. O período de expressão foi de 20h, rotação de 220 rpm e com a temperatura de 37 °C. A avaliação da expressão foi realizada em SDS-PAGE a 12% e *Western Blot* (Figuras 3 e 4).

[49] Para o teste de solubilidade o cultivo final foi centrifugado a 10.000 x g 10 min 4°C, e o sedimento foi ressuspenso com tampão NaP 50mM pH 8,0, com e sem ureia 6M. As bactérias foram lisadas utilizando sonicador, com três ciclos de 30 segundos na amplitude de 50 Hz. A avaliação da solubilidade dos antígenos recombinantes foi realizada em SDS-PAGE a 12% (Figura 5).

[50] Apenas o mutante 3 não foi expresso, mesmo realizando alterações nos parâmetros de produção. E quanto às proteínas expressas elas foram produzidas

tanto na fração solúvel como na insolúvel como foi observado no SDS-PAGE. Tal resultado é promissor já que a obtenção de proteínas recombinantes solúveis é um constante gargalo na área biotecnológica, além do que a solubilidade em tampões fisiológicos é parte fundamental da avaliação da viabilidade e eficiência do processo de produção.

Exemplo 3: Purificação do alérgeno Der p 21 tipo-selvagem e seus derivados

[51] A purificação das proteínas recombinantes apresentando cauda de 6XHis foi realizada por cromotografia de afinidade a íons de níquel. Para isso uma alíquota de 1 mL do cultivo foi centrifugada a 11.000 x g 2 min e em seguida foi utilizado o sistema de purificação de proteínas MagneHis ™ (Promega, Madison, WI, EUA), seguindo as orientações do fabricante de acordo com o resultado do teste de solubilidade (condição nativa ou desnaturante). O resultado da purificação foi analisado em SDS-PAGE 12% (Figura 3). A concentração proteica foi medida pelo *kit* Qubit™ Assay Protein Assay (Thermo Fisher Scientific, Waltham, MA, EUA).

Exemplo 4: Ensaios de Dot blot e ELISA para avaliar a reatividade de IgE com o Der p 21 wt e seus mutantes

[52] Para avaliar a reatividade de IgE com o Der p 21 tipo-selvagem e seus mutantes foram utilizados soros de indivíduos alérgicos e não alérgicos para o *Dermatophagoides pteronyssinus* obtidos da coorte do Programa de Controle da Asma e da Rinite Alérgica na Bahia (ProAR). O diagnóstico de alergia foi baseado em um estudo clínico de sintomas indicativos de alergia aos ácaros da poeira (rinite, conjuntivite e/ou asma), no teste cutâneo positivo para o ácaro *Dermatophagoides pteronyssinus*, e presença de IgE específica para esse ácaro > 0,7 kUA/L no soro do paciente, determinada com o ImmunoCAP.

[53] Para a detecção de IgE específica, através da técnica de *dot blot* foi utilizado um *pool* de soros de cinco indivíduos alérgicos ao *Dermatophagoides pteronyssinus*, e para o controle negativo foi utilizado um *pool* de soros de cinco indivíduos não alérgicos a esse ácaro. Inicialmente foi aplicado numa membrana de nitrocelulose (GE Healthcare, Waukesha, WI, USA) alíquotas de 2 µL, contendo 2,5 µg de cada

proteína recombinante purificada (Der p 21 tipo-selvagem e os mutantes 1, 2 e 4). O extrato do *Dermatophagoides pteronyssinus* e o alérgeno Der p 2 foram utilizados como controle positivo. Após o tempo de secagem da membrana (1,5h), ela foi bloqueada por 2 horas à TA (PBS 1X; 3% Tween-20 e 5% leite em pó desnatado). A incubação com o *pool* de soros dos indivíduos (alérgicos e não-alérgicos, diluição 1:5) foi feita à 4°C por 16 h. Em seguida a membrana foi incubada com o anticorpo anti-IgE humano biotinilado (BD Biosciences, San Jose, CA) (1:5000) por 1 hora na TA e com Estreptavidina-HRP (BD Biosciences, San Jose, CA) (1:2500) por 45 min. As bandas foram visualizadas com o substrato quimioluminescente ECL (GE Healthcare, Waukesha, WI, USA). A detecção do sinal quimioluminescente foi realizada no ImageQuant LAS 4000 (GE Healthcare, Waukesha, WI, USA). A análise densitométrica foi realizada a fim de quantificar o percentual de redução da ligação da IgE, para isso os *spots* foram analisados em um software de processamento e análise de imagens – ImageJ. No dot blot, observou-se que o mutante 1 apresentou a mais baixa porcentagem de ligação quando comparado com o alérgeno tipo-selvagem (Figura 6).

[54] Para outro ensaio de detecção de IgE alérgeno-específica (ELISA), utilizou-se placas de alta ligação (*Greiner* Bio-One MICROLON™) sensibilizadas com os antígenos (proteínas recombinantes purificada (Der p 21 tipo-selvagem e os mutantes 1, 2 e 4) na concentração de 5,0 µg/mL em tampão carbonato/bicarbonato pH 9,6–9,8, 12h a 4 °C. As placas foram bloqueadas por 12h a 4°C com PBS Tween-20 a 0,05% + 10% soro fetal bovino. Em seguida foram incubadas com o *pool* de soro de cinco indivíduos (alérgicos e não alérgicos ao *Dermatophagoides pteronyssinus*, diluídos 1:5 em TBS-T/BSA, por 16h a 4 °C. As placas foram incubadas com o conjugado anti-IgE humano biotinilado (BD Biosciences, San Jose, CA) (1:2000), TA e com Estreptavidina-HRP (BD Biosciences, San Jose, CA) (1:1000) ambos por 1h TA. A a revelação do ensaio foi realizada com o tampão de revelação (TMB 0,1 mg/mL e peróxido de hidrogênio 30%). A absorbância foi medida a 450 nm utilizando um leitor de placas (Multiskan™ FC Microplate Photometer). Os resultados do ELISA mostraram que quando comparadas com a molécula recombinante tipo-selvagem, todas as proteínas mutantes apresentaram redução na ligação da IgE, sendo que o mutante 1 apresentou a maior redução (Figura 7).

FIGURAS

Figura 1

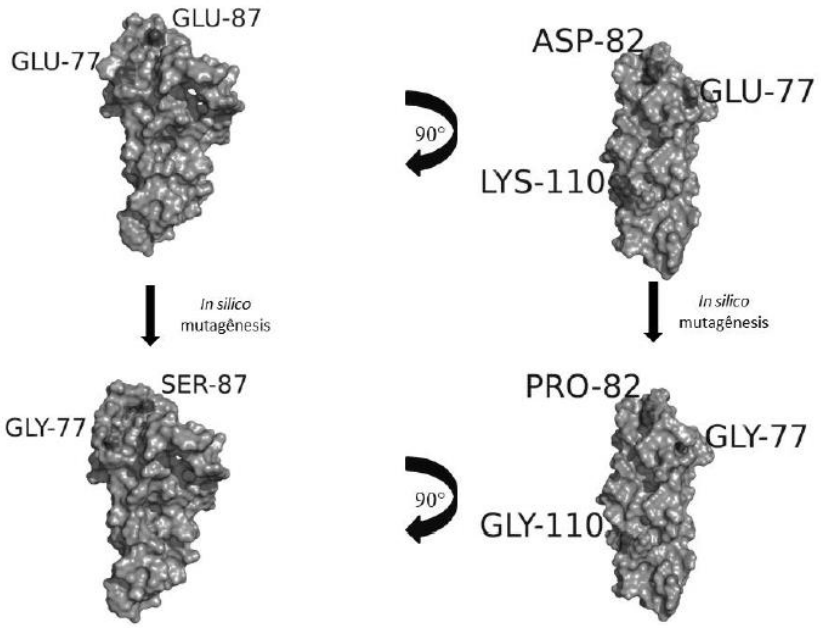


Figura 2

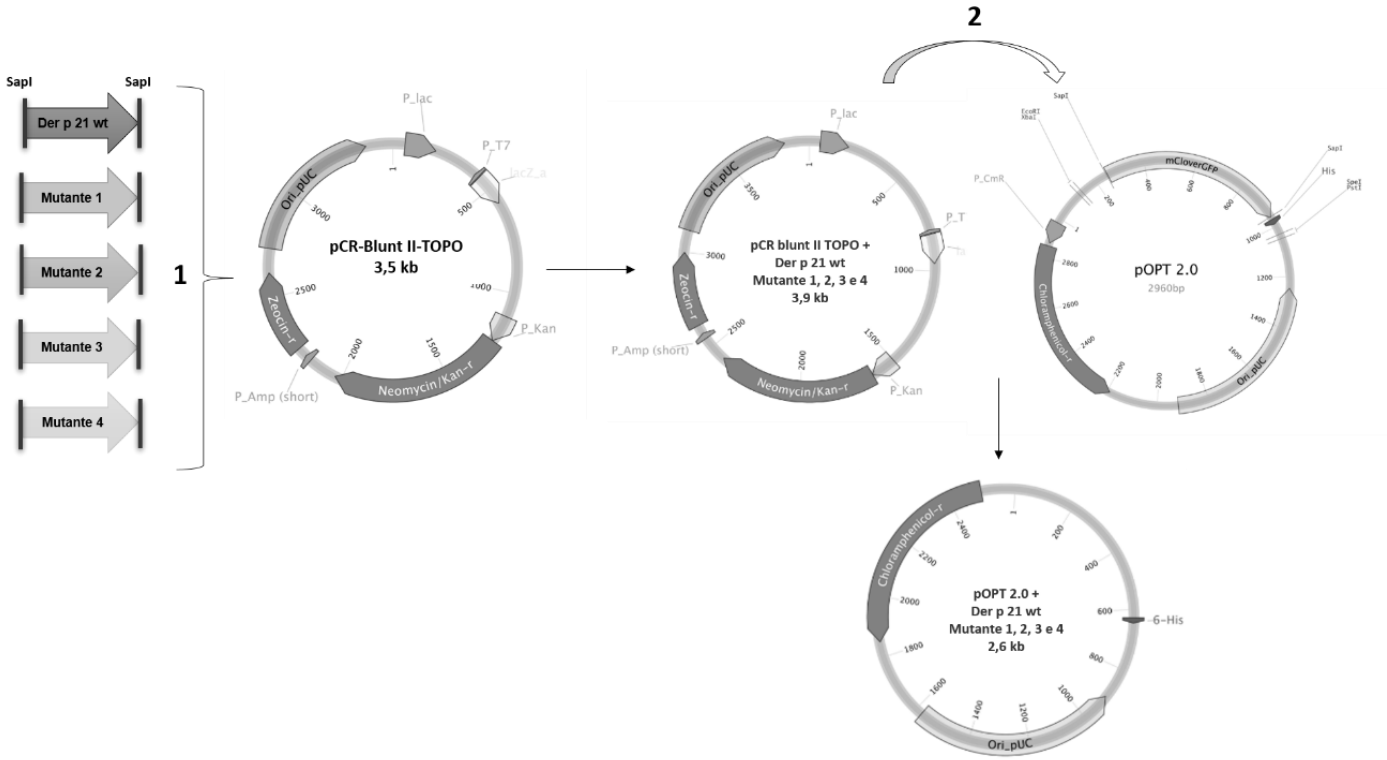

SapI
SapI
Der p 21 wt
Mutante 1
Mutante 2
Mutante 3
Mutante 4
1
pCR-Blunt II-TOPO
3,5 kb
P_lac
P_T7
P_Kan
P_Amp (short)
Ori_pUC
Zeocin-r
Neomycin/Kan-r
2
pCR blunt II TOPO +
Der p 21 wt
Mutante 1, 2, 3 e 4
3,9 kb
pOPT 2.0
2960bp
mCloverGFP
P_CmR
Chloramphenicol-r
pOPT 2.0 +
Der p 21 wt
Mutante 1, 2, 3 e 4
2,6 kb
6-His

Figura 3

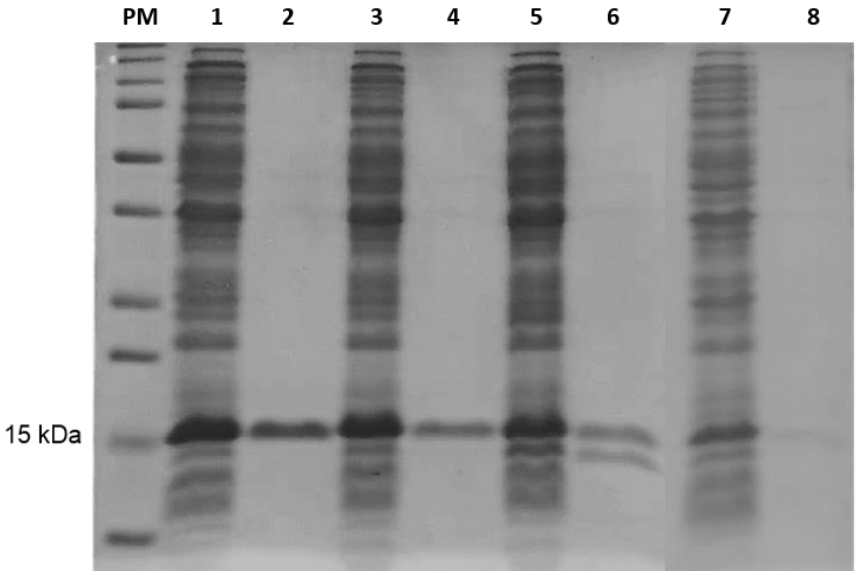


Figura 4

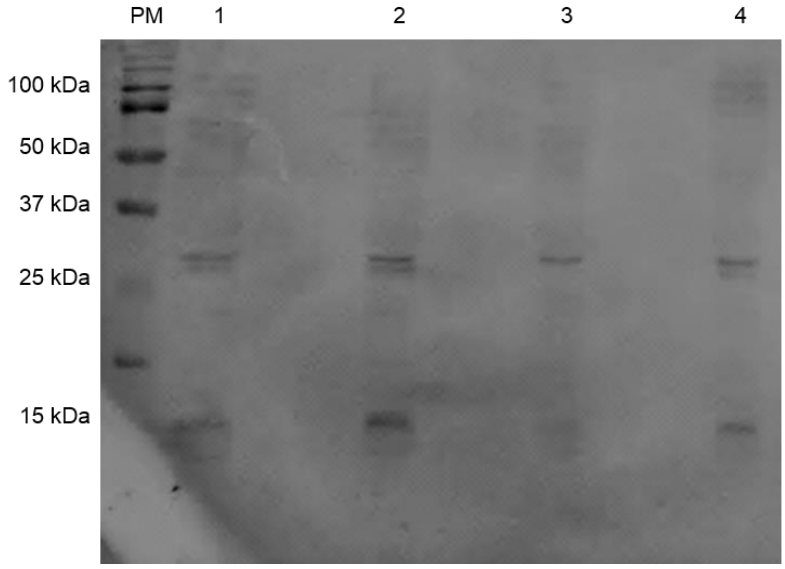


Figura 5

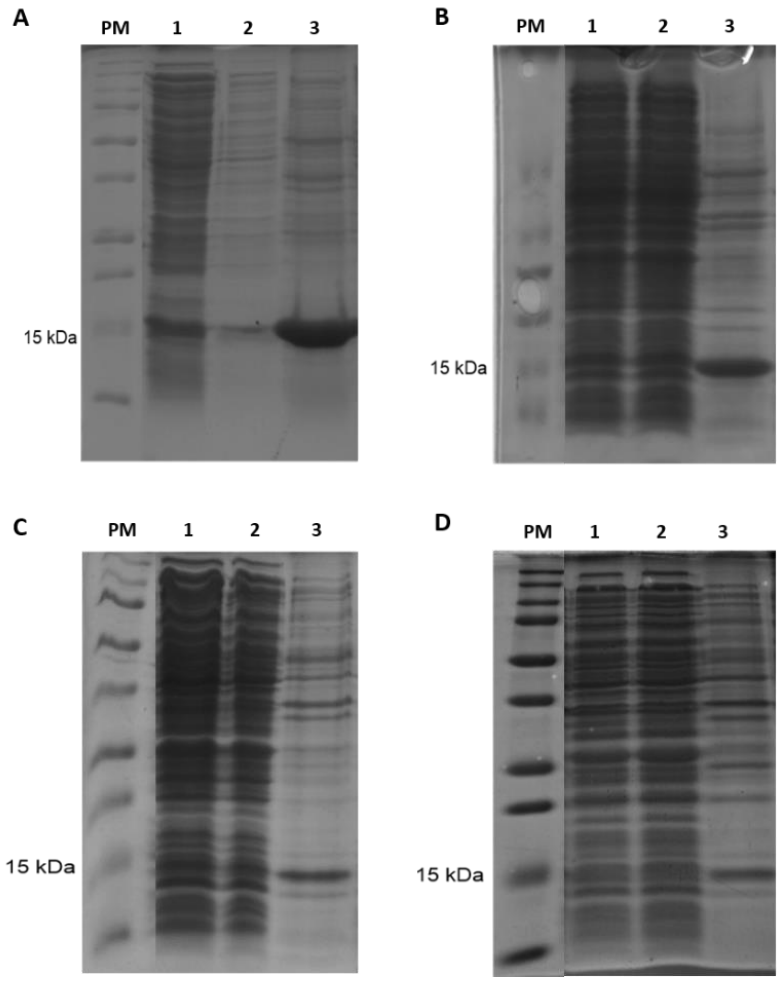


Figura 6

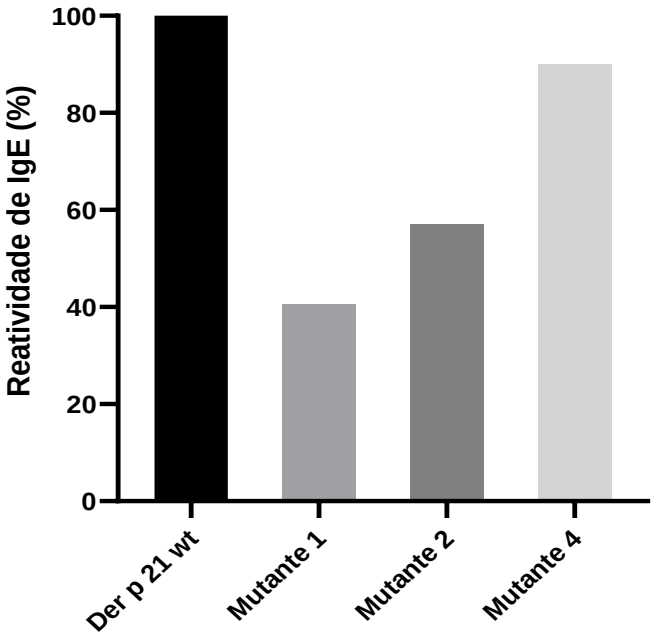
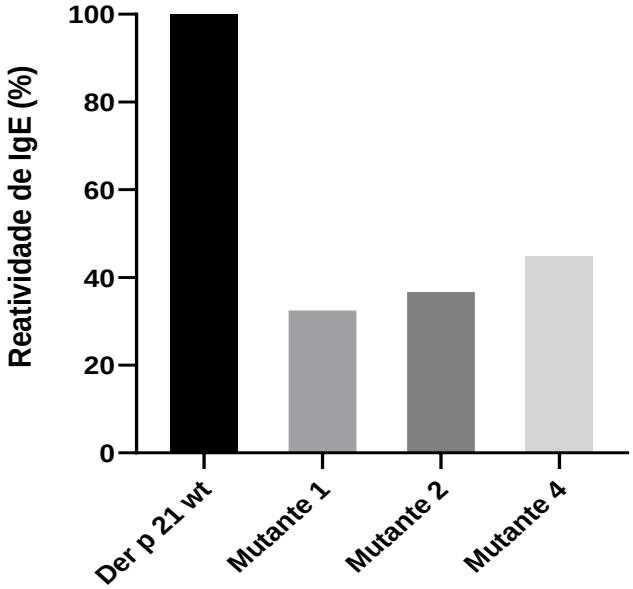


Figura 7



SEQUÊNCIAS

SEQ ID NO.: 1
FIVGDKKEDEWRMAFDRLMMEELETKIDQVEKGLLHLSEQYKELEKTKSKELKEQIL
RELTIGENFMKGALKFFEMEAKRTPLNMFERYNYEFALESIKLLIKKLDELAKKVKAV
NPDEYY

SEQ ID NO.: 2
FIVGDKKEDEWRMAFDRLMMEELETKIDQVEKGLLHLSEQYKELEKTKSKELKEQIL
RELTIGENFMKGALKFFEMEAKRTPLNMFERYNYEFALESIKLLIKKLDELAGKVKAV
NPDEYY

SEQ ID NO.: 3
FIVGDKKEDEWRMAFDRLMMEELETKIDQVEKGLLHLSEQYKELEKTKSKELKEQIL
RELTIGENFMKGALKFFEMGAKRTPLNMFERYNYEFALESIKLLIKKLDELAGKVKAV
NPDEYY

SEQ ID NO.: 4
FIVGDKKEDEWRMAFDRLMMEELETKIDQVEKGLLHLSEQYKELEKTKSKELKEQIL
RELTIGENFMKGALKFFEMGAKRTPLNMFSRYNYEFALESIKLLIKKLDELAGKVKA
VNPDEYY

SEQ ID NO.: 5
TTT ATT GTG GGG GAC AAA AAG GAA GAT GAG TGG CGT ATG GCG TTT GAC
CGG TTG ATG ATG GAA GAA CTG GAG ACT AAA ATC GAT CAG GTT GAG AAA G
GG CTG CTGCAC TTA TCG GAA CAA TAT AAG GAG CTG GAA AAA ACC AAA AG
T AAG GAA CTT AAA GAA CAG ATA TTG CGG GAA TTA ACT ATC GGT GAG AAC
TTC ATG AAA GGC GCATTA AAG TTT TTT GAG ATG GAG GCA AAA CGC ACC C
CA TTG AAT ATG TTT GAA CGG TAC AAT TAC GAG TTC GCA CTG GAG TCC ATC
AAA CTT TTG ATT AAG AAA TTAGAC GAA CTT GCA AAA AAA GTA AAA GCA GTT
AAT CCA GAC GAG TAC TAC

SEQ ID NO.: 6
TTT ATT GTA GGC GAT AAA AAA GAG GAT GAG TGG CGC ATG GCT TTC GAT A
GA CTT ATG ATG GAG GAG TTG GAG ACG AAG ATT GAC CAA GTT GAG AAG G
GT CTG CTGCAT CTG TCC GAG CAG TAC AAG GAG CTG GAG AAA ACG AAA T
CC AAG GAA TTA AAG GAA CAA ATC CTG CGT GAA CTG ACC ATC GGG GAG A
AC TTT ATG AAA GGA GCGTTA AAA TTT TTT GAA ATG GAA GCG AAG CGG ACG
CCC TTG AAT ATG TTC GAG CGT TAC AAC TAT GAG TTT GCC CTG GAA TCG A
TA AAA CTG CTG ATA AAG AAA TTAGAC GAA TTG GCT GGC AAA GTC AAA GCT
GTG AAT CCA GAC GAG TAT TAC

SEQ ID NO.: 7
TTC ATT GTC GGA GAT AAG AAA GAG GAC GAG TGG CGT ATG GCC TTT GAT
CGG CTT ATG ATG GAA GAG CTG GAG ACA AAG ATC GAT CAG GTA GAAAAA G
GG TTA TTG CAT TTG TCC GAG CAG TAC AAG GAA CTG GAG AAG ACG AAA A
GT AAA GAA CTG AAA GAA CAA ATA CTG CGG GAA TTA ACC ATTGGA GAG AAT
TTC ATG AAA GGA GCC TTA AAG TTC TTT GAG ATG GGC GCT AAA CGC ACC C
CG TTA AAC ATG TTC GAA CGC TAC AAT TAC GAA TTTGCG CTG GAA AGC ATA
AAG CTT TTA ATC AAG AAA CTG GAC GAG TTG GCT GGT AAA GTG AAA GCT G

TG AAT CCC GAT GAG TAC TAC

SEQ ID NO.: 8

TTT ATC GTC GGC GAT AAG AAG GAA GAC GAA TGG CGG ATG GCG TTC GAT CGC TTA ATG ATG GAA GAA TTG GAA ACC AAG ATA GAT CAA GTT GAA AAA GG T TTG CTGCAC CTG AGT GAA CAG TAC AAG GAA TTG GAA AAG ACC AAG AGT AAA GAG CTG AAG GAG CAA ATA TTG CGG GAA TTA ACG ATA GGG GAG AAT T TT ATG AAG GGC GCTTTA AAG TTT TTT GAG ATG GGG GCT AAG CGG ACG CC C TTA AAT ATG TTT TCC CGC TAT AAT TAC GAA TTT GCA CTG GAA TCC ATC AA G CTG CTG ATT AAA AAG CTGGAC GAG TTA GCA GGT AAG GTA AAG GCT GTG AAT CCG GAC GAA TAC TAT

REIVINDICAÇÕES

1. PROTEÍNAS RECOMBINANTES HIPOALERGÊNICAS *caracterizadas por* proteínas recombinantes derivadas de mutações na sequência de aminoácidos do alérgeno Der p 21 do ácaro da poeira doméstica *Dermatophagoides pteronyssinus*, as quais exibem reduzida atividade alergênica, possibilitando seu uso na profilaxia e tratamento de doenças alérgicas;

2. PROTEÍNAS RECOMBINANTES HIPOALERGÊNICAS de acordo com a reivindicação 001 *caracterizadas por* uma sequência de aminoácidos que deriva de modificações específicas *in silico* na sequência codificadora existente para o alérgeno Der p 21 tipo-selvagem, o qual ainda não possui função biológica conhecida, mas apresenta alta capacidade de ligação à IgE e caracterizado com peso molecular de aproximadamente 15 kDa e com cerca de 122 resíduos de aminoácidos;

3. PROTEÍNAS RECOMBINANTES HIPOALERGÊNICAS de acordo com as reivindicações 001 e 002 *caracterizadas por* sequências de aminoácidos que possuem mutações específicas *in silico* responsáveis pela redução na reatividade de IgE, quando comparadas com o alérgeno Der p 21 tipo-selvagem, representadas pelas sequências de aminoácidos: SEQ ID N°.: 1, SEQ ID N°.: 2, SEQ ID N°.: 3 e SEQ ID N°.: 4;

4. PROCESSO DE PRODUÇÃO DAS PROTEÍNAS RECOMBINANTES HIPOALERGÊNICAS de acordo com reivindicações 001 a 003 *caracterizado por* compreender as seguintes etapas:

- Busca da sequência do alérgeno Der p 21 em bancos de dados;
- Análise computacional de domínios e de estruturas secundárias;
- Análise computacional de sítios de modificações pós-traducionais;
- Análise computacional para predição dos epítopos de células B;
- Seleção de aminoácidos com escores mais altos de acordo os resultados de predição dos epítopos de células B para mutagênese *in silico*, especificamente, foram selecionadas as posições 77 – ácido glutâmico, 82 – ácido aspártico, 87 – ácido glutâmico, 110 – lisina;

- Varredura computacional de mutações desestabilizadoras nestas posições de aminoácidos, as 4 mutações mais desestabilizadoras (E77G, D82P, E87S e K110G) foram utilizadas para a produção das variantes mutantes do Der p 21 *wt* (Mutante 1, 2, 3 e 4);
- Realização de uma filtragem bioquímica para que não fossem selecionadas mutações em locais que causassem uma ruptura na estrutura secundária ou terciária da proteína;
- Análise das mutações na molécula tridimensional por mutagênese *in sílico*;
- Síntese química das sequências gênicas codificadoras das proteínas mutantes e clonagem em vetor plasmidiano de expressão;

5. MÉTODO de acordo com reivindicação 004 *caracterizado por* construção *in silico* de proteínas hipoalergênicas usando como molde a sequência de nucleotídeos que codifica o alérgeno Der p 21 tipo-selvagem, a referida construção *in silico* envolve a substituição de até quatro resíduos de aminoácidos da molécula tipo-selvagem que leva à alteração da superfície da estrutura da proteína original, com a finalidade de reduzir a capacidade de ligação à IgE, originando as sequências codificadoras otimizadas identificadas como SEQ ID N°.: 5, SEQ ID N°.: 6, SEQ ID N°.: 7 e SEQ ID N°.: 8;

6. SEQUÊNCIAS CODIFICADORAS DAS PROTEÍNAS RECOMBINANTES HIPOLAERGÊNICAS de acordo com reivindicações 004 e 005 *caracterizadas por* presença de mutações específicas em relação à sequência codificadora do alérgeno Der p 21 tipo-selvagem, além de outras versões destas sequências: reversas ou complementares, e contendo mutações semelhantes derivadas de processos de otimização e/ou harmonização de códons para expressão recombinante;

7. MÉTODO de acordo com reivindicações 004 a 006 *caracterizado por* o desenho do vetor de expressão contendo a unidade operacional de transcrição que incluem as sequências de nucleotídeos SEQ ID N°.: 5, SEQ ID N°.: 6, SEQ ID N°.: 7 e SEQ ID N°.: 8, bem como suas variantes de acordo com a reivindicação 006, para ser possível a expressão da versão recombinante das proteínas hipoalergênicas;

8. MÉTODO de acordo com reivindicações 004 a 007 *caracterizado por* produção heteróloga das proteínas hipoalergênicas em células procariotas, eucariotas e/ou sistemas *cell-free*; abrange-se sistemas de expressão recombinante que envolvam a fusão dessas proteínas hipoalergênicas das SEQ ID N°.: 1, SEQ ID N°.: 2, SEQ ID N°.: 3 e SEQ ID N°.: 4 com outras proteínas e/ou *tags*, bem como mudanças das sequências nucleotídicas SEQ ID N°.: 5, SEQ ID N°.: 6, SEQ ID N°.: 7 e SEQ ID N°.: 8, com a finalidade de purificação e/ou otimização do processo de obtenção da versão recombinante das proteínas hipoalergênicas, conforme solicitado nas reivindicações 001 a 003;

9. UMA VACINA PARA A IMUNOTERAPIA ALÉRGENO-ESPECÍFICA *caracterizada por* utilização das proteínas hipoalergênicas na sua formulação, conforme solicitado nas reivindicações 001 a 003, apresentando o hipoalérgeno na sua forma purificada ou combinado a agentes excipientes e/ou adjuvantes;

10. COMPOSIÇÃO VACINAL PARA A IMUNOTERAPIA ALÉRGENO-ESPECÍFICA de acordo com a reivindicação 009 caracterizada por serem formuladas na forma de uma solução, suspensão, emulsão, liofilizada ou sistema transdérmico;

11. COMPOSIÇÃO VACINAL PARA A IMUNOTERAPIA ALÉRGENO-ESPECÍFICA de acordo com a reivindicação 009 caracterizada por administração via subcutânea, sublingual, bucal, nasal, retal, tópica, por inalação, ou administração parentérica.

8.2 TEXTO COMPLETO DOS ARTIGOS PUBLICADOS DURANTE O DOUTORADO

Título	**Revista**	**DOI**	**Data de publicação**	**Autoria**
Engineering an Optimized Expression Operating Unit for Improved Recombinant Protein Production in *Escherichia coli*	Protein Expression and Purification	10.1016/j.pep.2022.106150	06/08/2022	Primeiro autor
Rationally designed hypoallergenic mutant variants of the house dust mite allergen Der p 21	Biochimica et Biophysica Acta (BBA) - General Subjects	10.1016/j.bbagen.2022.130096	22/01/2022	Primeiro autor
In vivo cleavage of solubility tags as a tool to enhance the levels of soluble recombinant proteins in *Escherichia coli*	Biotechnology and Bioengineering	10.1002/bit.27912	9/08/2021	Co-autor
Immunogenicity and protection induced by recombinant *Toxocara canis* proteins in a murine model of toxocariasis	Vaccine	10.1016/j.vaccine.2020.04.072	19/06/2020	Co-autor

Single-Input Regulatory Cascade for in vivo Removal of the Solubility Tag in Fusion Recombinant Proteins Produced by *Escherichia coli*	Frontiers in Bioengineering and Biotechnology	10.3389/fbioe .2019.00200	20/08/2019	Co-autor

Protein Expression and Purification 199 (2022) 106150

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Engineering an optimized expression operating unit for improved recombinant protein production in *Escherichia coli*

Sara P.O. Santos [a], Luis Fabian S. Garcés [a,1], Filipe S.R. Silva [a,1], Leonardo F. Santiago [a], Carina S. Pinheiro [a], Neuza M. Alcantara-Neves [a], Luis G.C. Pacheco [a,*]

[a] Institute of Health Sciences, Federal University of Bahia, Salvador, BA, Brazil

ARTICLE INFO

Keywords:
Escherichia coli
Expression operating unit
Recombinant protein
Synthetic biology

ABSTRACT

Common strategies to improve recombinant protein production in *Escherichia coli* often involve the test and optimization of several different variables, when using traditional expression vectors that are commercially available. Now, modern synthetic biology-based strategies allow for extensive modifications of these traditional vectors, or even construction of entirely new modular vectors, so as to permit tunable production of the recombinant proteins of interest. Herein, we describe the engineering of a new expression operating unit (EOU; 938 bp) for producing recombinant proteins in *E. coli*, through the combinatorial assembly of standardized and well-characterized genetic elements required for transcription and translation (promoter, operator site, RBS, junction RBS-CDS, cloning module, transcriptional terminator). We also constructed a novel T7 promoter variant with increased transcriptional activity (1.7-fold higher), when compared to the canonical wild type T7 promoter sequence. This new EOU yielded an improved production of the reporter protein superfolder GFP (sfGFP) in *E. coli* BL21(DE3) (relative fluorescence units/RFU = 70.62 ± 1.62 A U.) when compared to a high-producing control expression vector (plasmid BBa_I746909; RFU = 59.68 ± 1.82 A U.). The yields of purified soluble recombinant sfGFP were also higher when using the new EOU (188 mg L^{-1} culture vs. 108 mg L^{-1} in the control) and it performed similarly well when inserted into different plasmid backbones (pOPT1.0/Amp^R and pOPT2.0/Cm^R).

1. Introduction

Recombinant protein production in bacteria is essential for basic research in the life sciences and for applications in the therapeutic and industrial fields [1–6]. *Escherichia coli* is the microbial workhorse for the production of heterologous proteins and remains as the most widely used expression host [1,3,5,7–13]. It is estimated that *ca.* 88% of protein structures deposited in the Protein Data Bank are derived from proteins produced in this host organism [12]. In addition, *E. coli* is utilized as a host in about 75% of the academic studies involving recombinant proteins [14], and also in the production of nearly 30% of the recombinant therapeutic biomolecules by the biotechnology industry [12].

To circumvent difficulties commonly associated with recombinant protein production in bacteria, such as low protein yield and aggregation, different strategies are often employed that involve optimization of culturing conditions, and protocols to improve protein solubility and purification [4,15–18]. Additionally, several variables are traditionally optimized to improve heterologous protein production in *E. coli*, including the test of multiple genetically engineered host strains and the selection of the most appropriate vector among a range of commercially available expression plasmids, such as the families pET, pBAD and pTrc [19–21].

However, modern strategies to produce recombinant proteins in *E. coli* have involved either modifications in these traditionally used expression vectors [22,23] or the construction of entirely new modular vectors so as to permit tunable production of the proteins of interest [3,21,24]. These strategies rely on recent advances in the synthetic biology field, that allow for better design, selection, synthesis, combination, and systematic analysis of new genetic elements, then improving expression vectors and recombinant protein production [25]. This synthetic biology-guided construction of improved genetic elements has already been evaluated in the optimization of promoter elements [3,26–28],

* Corresponding author.Department of Biotechnology
E-mail address: luis.pacheco@ufba.br (L.G.C. Pacheco).
[1] Present addresses: L.F·S.G.: Technical University of Ambato, Ecuador. F.S.R.S.: University of São Paulo, São Carlos-SP, Brazil.

https://doi.org/10.1016/j.pep.2022.106150
Received 27 June 2022; Received in revised form 26 July 2022; Accepted 27 July 2022
Available online 6 August 2022

regulatory operator sites [29], ribosome binding sites (RBS) [30–33], coding sequences (CDS) [25,34–37], and of transcription/translation initiation regions, such as the junctions between untranslated 5′ region (5′-UTR)/RBS/CDS [38–40]. All these studies described extensively characterized and standardized biological parts that can now be easily reused to construct new improved genetic tools [41–43].

In this study, we constructed a novel Expression Operating Unit (EOU) for recombinant protein expression in *E. coli* based on the combination of several well-characterized genetic elements for improved transcription and translation of a gene of interest (Fig. 1A). Additionally, a new optimized T7 promoter variant was constructed that combined sequence alterations previously shown to improve transcription by separate studies (Fig. 1B) [26,27]. This new EOU was cloned in different plasmid backbones to generate the expression vectors pOPT1.0 and pOPT2.0, whose efficiencies to express a protein of interest were then evaluated using the superfolder green fluorescent protein (sfGFP) as a reporter.

2. Materials and methods

2.1. Selection of standardized genetic elements and design of the Expression Operating Unit (EOU)

The construction of the new Expression Operating Unit (EOU) was based on tried-and-tested standardized genetic parts retrieved from the Registry of Standard Biological Parts (RSBP, http://parts.igem.org/Main_Page) or from selected studies through literature search (Supplementary Table S1). The constructed EOU (Fig. 1A) is driven by a modified T7 promoter variant, which carries a combination of mutations previously shown to render increased transcription efficiency [26,27]. Basically, four nucleotides were mutated in the internal sequence of the wild-type T7 promoter, the extension CCGGT was inserted in the 5′ end, and the 3’ terminus remained unchanged (Fig. 1B).

Besides the modified T7 promoter variant, the EOU was composed of: two consecutive lac operators (*lacOs* and *lacO3*), such as in Ref. [29]; a strong ribosome binding site (RBS) (BBa_B0034); a well-defined RBS-CDS junction sequence, as described in Ref. [38]; a cloning module containing the coding sequence for the superfolder GFP reporter protein (*sfgfp* – 720 bp) flanked by restriction sites for the type IIS enzyme *SapI*, for streamlined CDS replacement; a region coding for an in-frame 6x histidine tag; and a T7 terminator site (BBa_B1006). To standardize the cloning steps, the prefix and suffix sequences for the RFC23 cloning standard were added to the 5′ and 3′ ends of the EOU, respectively [44] (Fig. 1C; Supplementary Fig. S1).

The EOU (938 bp) was firstly inserted in pUC57, to yield the vector pOPT 1.0 (3.65 Kbp; pUC_ori/Amp^R) (Supplementary Fig. S2). The EOU was also sub-cloned in the pSB1C3 backbone, through digestion with *EcoRI* and *PstI* to generate the vector pOPT 2.0 (Supplementary Fig. S2; Supplementary Table S2).

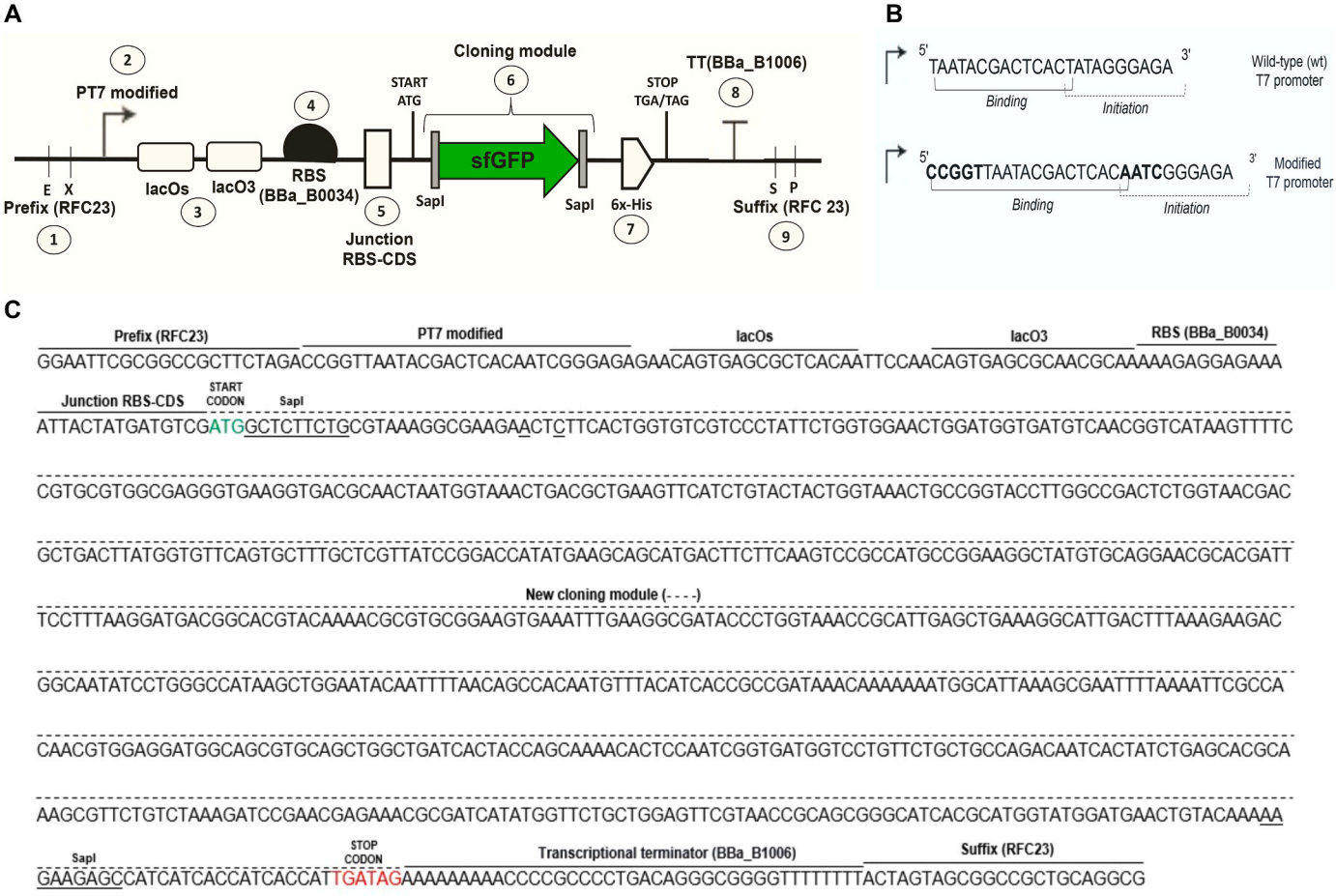


Fig. 1. Schematic of the engineered standardized Expression Operating Unity (EOU). (A) Standard genetic elements that compose the EOU: 1- standard prefix RFC 23; 2- modified T7 promoter variant; 3- operator sites (*lacOs* and *lacO3*); 4- strong RBS (BBa_B0034); 5- RBS-CDS junction; 6- Type IIS-based cloning module; 7- 6xHis C-terminal tag; 8- transcriptional terminator (BBa_B1006); 9- standard suffix RFC 23. (B) Sequence comparison between the modified T7 promoter variant and the wild-type T7 promoter. Alterations are shown in bold. (C) Sequence map of the EOU.

2.2. Protein expression standardization using the new EOU

The protein sfGFP was used as a reporter to standardize the recombinant protein expression from the pOPT 1.0 vector transformed in different *E. coli* strains (Supplementary Table S3). Measurement of sfGFP fluorescence in the culture was initially used as an indicative of total protein production. The plasmid BBa_I746909, obtained from the iGEM Registry of Standard Biological Parts (http://parts.igem.org/Part:BBa_I746909), was used as a positive control. This plasmid carries the coding sequence for the sfGFP protein driven by the canonical wild-type T7 promoter, a strong RBS sequence (BBa_B0034), and two transcriptional terminators (B0010 and B0012) (Supplementary Table S2; Supplementary Fig. S3). Fluorescence from *E. coli* cultures was monitored by Fluoroskan Ascent Microplate Fluorometer (Thermo Fisher) (Ex. = 485 nm/Em = 538 nm).

2.3. Recombinant protein purification and analysis

To effectively quantify recombinant protein production, his-tagged sfGFP was purified from induced *E. coli* cultures. Firstly, bacterial extracts were analyzed by 12% SDS-PAGE and Western blot (mouse anti-His IgG, 1:4000). Protein solubility was evaluated by a standard protocol, as previously described [45]. The MagneHis Protein Purification System (Promega) was used for affinity purification of sfGFP. Total yields of recombinant sfGFP were quantified with the Qubit Protein Assay Kit (Thermo Fisher).

2.4. Transcriptional activity from the modified T7 promoter variant

Reverse-transcription quantitative real-time PCR was used to evaluate the transcriptional activity arising from the modified T7 promoter variant, in comparison to the wild-type T7 promoter. For this, we first selected from the iGEM Parts Registry a positive control plasmid (BBa_K567018) that renders a well-characterized and robust expression of a gene of interest (*egfp*) under the control of the canonical wild-type T7 promoter. We then constructed the pT7var plasmid (Supplementary Table S2) carrying an *egfp* expression cassette that has the same genetic structure from BBa_K567018 but changing the wt T7 promoter by the new variant. Both plasmids were transformed in *E. coli* BL21 (DE3) and *egfp* transcription was stimulated with 1 mM IPTG, at OD600nm = 0.3. Total RNA was isolated and then used in RT-qPCR experiments targeting *egfp* (Supplementary Methods).

2.5. Statistical analyses

GraphPad Prism software was used to perform statistical analyses. Normality of tested by the Kolmogorov-Smirnov and Shapiro-Wilk tests. ANOVA was used for variance analysis, with Bonferroni post-hoc test (95% significance level).

3. Results and discussion

Research efforts for enhancing the expression levels of recombinant proteins in bacterial hosts have been crucial for biomedical and biotechnological studies [4,5]. Such efforts now involve the application of synthetic biology strategies to the development of standardized and well-characterized genetic elements, so-called biological parts, aiming at improved transcriptional and translational rates of genes of interest in bacteria [3,21,42,43,46]. Herein, we assembled an optimized Expression Operating Unit (EOU) for recombinant protein production in *Escherichia coli* through reuse of some of these well-characterized biological parts that are currently available (Fig. 1A and C). This strategy of rationally designing new EOUs by leveraging extensively characterized genetic parts has already been shown to be an interesting alternative to avoid unnecessary trial and error work when trying to improve expression of a given protein of interest in bacteria [42,43,47].

Additionally, we created a new T7 promoter variant with enhanced transcriptional activity (Fig. 1B) through combination of previously tried and tested sequence changes [26,27]. Particularly, Paul and collaborators [26] created and screened a randomized T7 promoter library and selected mutants with high *in vitro* transcription activities. This study reported a T7 variant possessing substitutions towards the 3′-end of the sequence, which was able to improve the recombinant protein yield after translation. The study by Chizzolini and collaborators [27], in turn, achieved enhanced transcriptional activity from the wt T7 promoter through a 5 bp sequence extension (CCGGT) in the 5′-end. In our study, the combination of both sequence changes to form a modified T7 promoter variant (Fig. 1B) resulted in an additional 1.7-fold increase in transcriptional activity, when comparing by RT-qPCR the expression of the *egfp* gene from the plasmids pT7var and BBa_K567018 (Supplementary Table S4); this improvement was also translated into higher yields of purified sfGFP. Interestingly, in a recent study by Shilling and collaborators [23], it was reported that the truncated T7 promoter sequence present in the family of vectors of the pET expression system renders a three-fold reduction in the expression of the sfGFP protein, when compared to a consensus wild-type T7 promoter, which is four bases longer. The study suggests that the truncation in the T7 sequence found in pET arises from a design flaw that reduces protein production [23]. In our study, that was developed concomitantly, we extended this knowledge by demonstrating that the addition of tried and tested sequence extensions to the consensus T7 promoter sequence can further improve the expression of a protein of interest in *E. coli*.

Different studies [48,49] described the importance of using a T7 expression system with genetic elements that attenuate the toxicity generated by the overexpression that can be caused by this system. In this way, we created an expression vector with elements so that the expression was highly controlled. We used two consecutive lac operators (*lacOs* and *lacO3*) [29], which together have a higher affinity for the *LacI* repressor protein, and a transcriptional terminator (BBa_B1006) with high transcription termination rate (~90%).

Gustafsson and collaborators [25] took a more general approach and tested different experimental variables to improve the yield of recombinant proteins (bacterial strain, temperature, aeration, nutrients, pH). They also considered the genetic elements that impact the recombinant protein production, including the coding sequence (CDS) itself and the genetic elements that compose the expression vector. Following a similar reasoning, we tested different variables to find the ideal culture conditions to maximize the yield of the sfGFP recombinant protein produced from the new EOU (Supplementary Table S3). The best sfGFP (~28 kDa) expression in our study was achieved in the *E. coli* BL21 (DE3) strain grown in Luria-Bertani (LB), at 37 °C for 20 h, with agitation at 200 rpm and induction by 1 mM of IPTG at OD_{600nm}=0.3–0.6 (Fig. 2A and B). Most of the recombinant protein was produced in the soluble (Sol.) fraction after 20 h of induction, with only trace amounts detected at the insoluble (Insol.) fraction (Fig. 2A), even by Western blot analysis (Fig. 2B). The fluorescence emitted by the presence of soluble sfGFP in the bacterial cell extracts also corroborated the results observed by SDS-PAGE analysis (Fig. 2A, lower panel).

We conducted the experiments under the same conditions with the pOPT 1.0 plasmid, harboring the new EOU, and the positive control plasmid BBa_I746909. Then, we registered the fluorescence emission by the cell cultures over time, following induction (Fig. 2C; Supplementary Fig. S4). After 20 h, there was a significantly higher sfGFP production from cultures carrying the new EOU (mean sfGFP fluorescence = 70.62 ± 1.62 A U.), when compared to the positive control plasmid (mean sfGFP fluorescence = 59.68 ± 1.82 A U.) (Fig. 2C; Supplementary Fig. S5). This result was also reproducible when the EOU was sub-cloned in a different plasmid backbone, to create pOPT2.0 (Fig. 2D). The construction of pOPT2.0 was also important to permit the subsequent cloning of different target genes in the EOU, as we identified an additional *SapI* restriction site outside the EOU in the pOPT1.0 backbone when attempting to digest that plasmid.

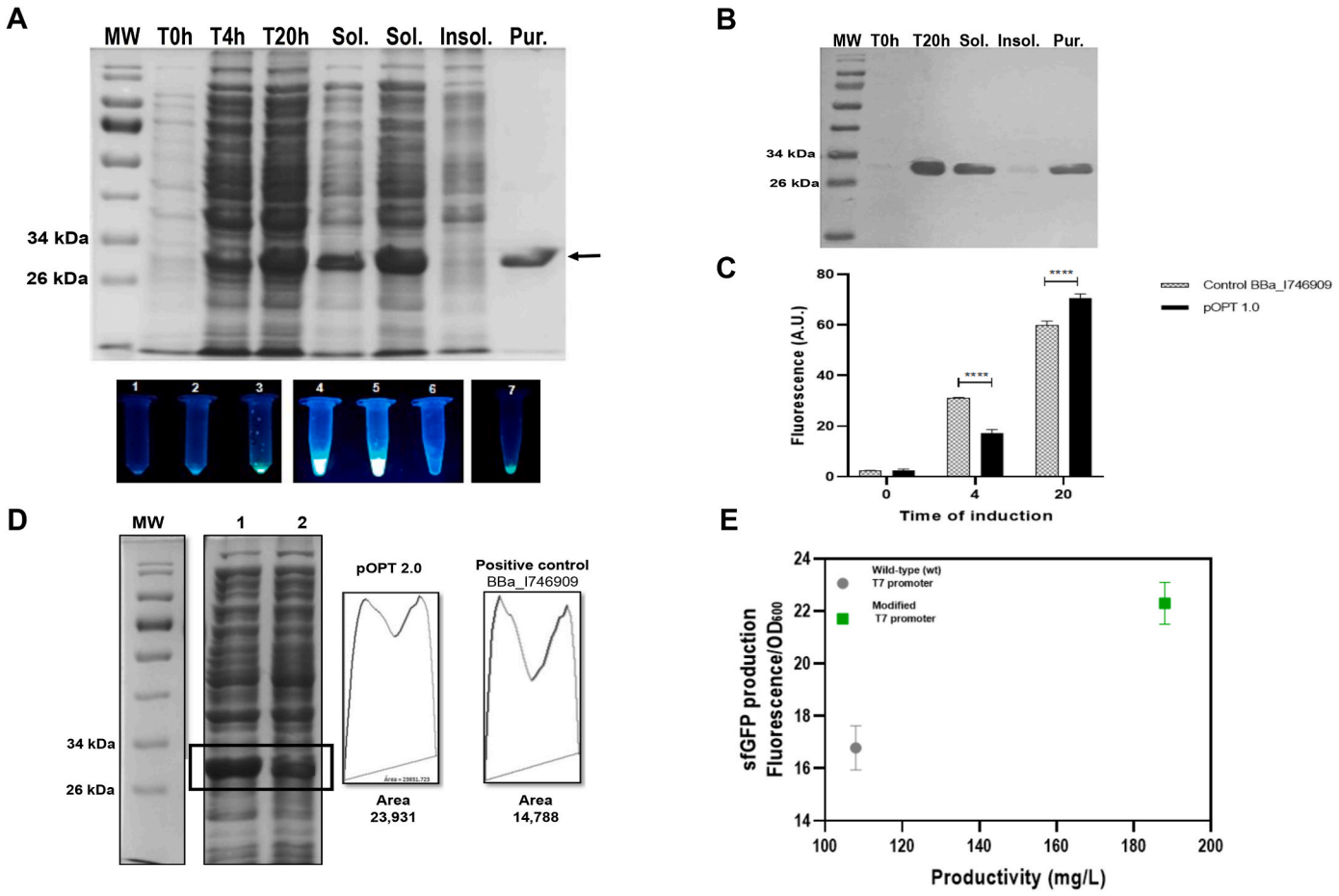


Fig. 2. Functioning of the new EOU evaluated through expression of recombinant sfGFP. (A) Upper panel, 12% SDS-PAGE showing expression of the sfGFP protein (ca. 28 kDa) in *E. coli* BL21 (DE3) transformed with the pOPT 1.0 vector. Time points (0 h, 4 h, 20 h) represent period of induction by IPTG. Sol. = soluble fraction; Insol. = insoluble fraction. Pur. = purified proteins. Lower panel, sfGFP fluorescence emitted by the *E. coli* cell extracts at each matched sample in the SDS-PAGE: 1- T0h; 2- T4h; 3- T20h; 4- Sol.1; 5- Sol.2; 6- Insol.; 7- Pur. (B) Western blot analysis of sfGFP expression in the BL21 (DE3) strain transformed with the pOPT1.0 vector. (C) sfGFP fluorescence from BL21 (DE3) cultures transformed with either the pOPT 1.0 plasmid or the positive control plasmid BBa_I746909, after IPTG induction. Data is presented as mean ± S.D. (n = 5). **** Significantly different ($p < 0.0001$). (D) SDS-PAGE and densitometric analysis by ImageJ of the soluble recombinant sfGFP produced by *E. coli* transformed with the pOPT 2.0 plasmid or the positive control BBa_I746909. (E) Yields of soluble recombinant sfGFP in *E. coli* BL21 (DE3) transformed with the pOPT 1.0 (modified P_{T7}) or the positive control BBa_I746909 (wild-type P_{T7}). sfGFP was purified to completion by affinity chromatography and the final yields were plotted against the sfGFP fluorescence of the culture normalized by the optical density at 600 nm.

Recombinant sfGFP was purified to completion by affinity chromatography (Fig. 2A and B), rendering final soluble protein yields of 188 mg L^{-1} culture from pOPT 1.0 and 108 mg L^{-1} culture from the positive control BBa_I746909 (Fig. 2E).

Producing high levels of recombinant proteins in bacteria in their soluble form can be an infrequent outcome, assuming the contradictory dynamics between these two variables in the recombinant protein expression context [12,50]. Obtaining biologically active and well-folded proteins is a bottleneck in many recombinant protein processes, having a huge demand in many industries [51–53]. Then, developing expression systems that prioritize these qualities should be the starting point of any project aiming to produce proteins of biotechnological interest.

In another recent study by our group, we already explored the effectiveness of our newly engineered EOU to express a protein of biotechnological interest, the house dust mite allergen Der p 21 [45]. In previous attempts to express this protein from commercially available vectors, such as pD444, we couldn't achieve a successful expression in *E. coli,* even when different expression conditions were tested (data not shown). Noteworthy, after cloning in the pOPT2.0 plasmid, carrying the new EOU, the Der p 21 allergen and its mutant derivatives could be produced in *E. coli* at final soluble protein yields ranging from 55.6 to 60 mg.L-1 of culture [45]. The difficulties associated with production of soluble recombinant proteins from eukaryotic origin using traditional commercial vectors was also demonstrated by Pacheco and collaborators [54] when testing various target genes in distinct *E. coli* strains, using a pET expression vector (pET28a-LIC) [54]. Therefore, engineering new optimized expression vectors using well-characterized genetic elements may be a good alternative to tackle the problem of obtaining recombinant proteins in *E. coli* at desirable quantity and quality.

4. Conclusion

In this study we engineered a new expression operating unit (EOU) for producing recombinant proteins in *E. coli*, through the combination of standardized and well-characterized genetic elements. We also constructed a novel T7 promoter variant with increased transcriptional activity, when compared to the canonical wild type T7 promoter sequence. The new EOU yielded an improved production of a sfGFP in *E. coli*, mostly in the soluble form, and performed similarly well in different plasmid backbones. The ability to produce a hard-to-express protein from eukaryotic origin was already demonstrated by production of the Der p 21 allergen from house dust mite [45]. This study demonstrates the utility of reusing deeply characterized genetic

elements in new combinations to develop optimized expression devices capable of improving recombinant protein production in bacteria.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgements

S.P.O.S. was recipient of a PhD scholarship from CAPES foundation, Brazil. N.M.A.N, and L.G.C.P. are recipients of research fellowships from CNPq-Brazil. This study was partially funded by FAPESB, through research grant #PNE0007/2014.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pep.2022.106150.

References

[1] N.K. Tripathi, A. Shrivastava, Recent developments in bioprocessing of recombinant proteins: expression hosts and process development, Front. Bioeng. Biotechnol. 7 (2019), https://doi.org/10.3389/fbioe.2019.00420.
[2] J. Puetz, F.M. Wurm, Recombinant proteins for industrial versus pharmaceutical purposes: a review of process and pricing, Processes 7 (2019) 476, https://doi.org/10.3390/pr7080476.
[3] O. Firat Duzenli, S. Okay, Promoter engineering for the recombinant protein production in prokaryotic systems, AIMS Bioeng 7 (2020) 62–81, https://doi.org/10.3934/bioeng.2020007.
[4] G. Lozano Terol, J. Gallego-Jara, R.A. Sola Martínez, A. Martínez Vivancos, M. Cánovas Díaz, T. de Diego Puente, Impact of the expression system on recombinant protein production in Escherichia coli BL21, Front. Microbiol. 12 (2021) 1–12, https://doi.org/10.3389/fmicb.2021.682001.
[5] A.A. Tungekar, A. Castillo-Corujo, L.W. Ruddock, So you want to express your protein in Escherichia coli? Essays Biochem. 65 (2021) 247–260, https://doi.org/10.1042/EBC20200170.
[6] B.K. Bhandari, C.S. Lim, P.P. Gardner, TISIGNER.com: web services for improving recombinant protein production, Nucleic Acids Res. 49 (2021) W654–W661, https://doi.org/10.1093/nar/gkab175.
[7] R. Chen, Bacterial expression systems for recombinant protein production: E. coli and beyond, Biotechnol. Adv. 30 (2012) 1102–1107, https://doi.org/10.1016/j.biotechadv.2011.09.013.
[8] L. Marschall, P. Sagmeister, C. Herwig, Tunable recombinant protein expression in E. coli: enabler for continuous processing? Appl. Microbiol. Biotechnol. 100 (2016) 5719–5728, https://doi.org/10.1007/s00253-016-7550-4.
[9] G.L. Rosano, E.S. Morales, E.A. Ceccarelli, New tools for recombinant protein production in Escherichia coli: a 5-year update, Protein Sci. 28 (2019) 1412–1422, https://doi.org/10.1002/pro.3668.
[10] K.A.R. Packiam, R.N. Ramanan, C.W. Ooi, L. Krishnaswamy, B.T. Tey, Stepwise optimization of recombinant protein production in Escherichia coli utilizing computational and experimental approaches, Appl. Microbiol. Biotechnol. 104 (2020) 3253–3266, https://doi.org/10.1007/s00253-020-10454-w.
[11] M.R. Ki, S.P. Pack, Fusion tags to enhance heterologous protein expression, Appl. Microbiol. Biotechnol. 104 (2020) 2411–2425, https://doi.org/10.1007/s00253-020-10402-8.
[12] F.S.R. Silva, S.P.O. Santos, R. Meyer, E.S. Silva, C.S. Pinheiro, N.M. Alcantara-Neves, L.G.C. Pacheco, In vivo cleavage of solubility tags as a tool to enhance the levels of soluble recombinant proteins in Escherichia coli, Biotechnol. Bioeng. 118 (2021) 4159–4167, https://doi.org/10.1002/bit.27912.
[13] Z.J. Li, Z.X. Zhang, Y. Xu, T.Q. Shi, C. Ye, X.M. Sun, H. Huang, CRISPR-based construction of a BL21 (DE3)-Derived variant strain library to rapidly improve recombinant protein production, ACS Synth. Biol. 11 (2022) 343–352, https://doi.org/10.1021/acssynbio.1c00463.
[14] R.M. Bill, Playing catch-up with escherichia coli: using yeast to increase success rates in recombinant protein production experiments, Front. Microbiol. 5 (2014) 1–5, https://doi.org/10.3389/fmicb.2014.00085.
[15] T. Kondo, S. Yumura, Strategies for enhancing gene expression in Escherichia coli, Appl. Microbiol. Biotechnol. 104 (2020) 3825–3834, https://doi.org/10.1007/s00253-020-10430-4.
[16] B.K. Bhandari, P.P. Gardner, TISIGNER . com : web services for improving recombinant protein production 49 (2021) 654–661.
[17] S. Schlegel, P. Genevaux, J.W. de Gier, Isolating Escherichia coli strains for recombinant protein production, Cell. Mol. Life Sci. 74 (2017) 891–908, https://doi.org/10.1007/s00018-016-2371-2.
[18] R. Dana, Advances and innovations in recombinant protein expression technology, J. Adv. Pharm. Educ. Res. 8 (2018) 88–91.
[19] I. Ahmad, N. Nawaz, N.M. Darwesh, S. ur Rahman, M.Z. Mustafa, S.B. Khan, S. G. Patching, Overcoming challenges for amplified expression of recombinant proteins using Escherichia coli, Protein Expr, Purif 144 (2018) 12–18, https://doi.org/10.1016/j.pep.2017.11.005.
[20] J. Kaur, A. Kumar, J. Kaur, Strategies for optimization of heterologous protein expression in E. coli: roadblocks and reinforcements, Int. J. Biol. Macromol. 106 (2018) 803–822, https://doi.org/10.1016/j.ijbiomac.2017.08.080.
[21] Y. Zhang, H. Gao, X. Qi, S. Gao, S. Xu, H. Wang, Y. An, Efficient molecular biological manipulations with improved strategies based on novel Escherichia coli vectors, J. Agric. Food Chem. 69 (2021) 5086–5095, https://doi.org/10.1021/acs.jafc.1c00109.
[22] T. Sadeghian-Rizi, A. Ebrahimi, F. Moazzen, H. Yousefian, A. Jahanian-Najafabadi, Improvement of solubility and yield of recombinant protein expression in E. coli using a two-step system, Res. Pharm. Sci. 14 (2019) 400–407, https://doi.org/10.4103/1735-5362.268200.
[23] P.J. Shilling, K. Mirzadeh, A.J. Cumming, M. Widesheim, Z. Köck, D.O. Daley, Improved designs for pET expression plasmids increase protein production yield in Escherichia coli, Commun. Biol. 3 (2020), https://doi.org/10.1038/s42003-020-0939-8.
[24] D. Mikiewicz, A. Plucienniczak, A. Bierczynska-Krzysik, A. Skowronek, G. Wegrzyn, Novel expression vectors based on the pIGDM1 plasmid, Mol. Biotechnol. 61 (2019) 763–773, https://doi.org/10.1007/s12033-019-00201-6.
[25] C. Gustafsson, J. Minshull, S. Govindarajan, J. Ness, A. Villalobos, M. Welch, Engineering genes for predictable protein expression, Protein Expr. Purif. 83 (2012) 37–46, https://doi.org/10.1016/j.pep.2012.02.013.
[26] S. Paul, A. Stang, K. Lennartz, M. Tenbusch, K. Überla, Selection of a T7 promoter mutant with enhanced in vitro activity by a novel multi-copy bead display approach for in vitro evolution, Nucleic Acids Res. 41 (2013) 1–11, https://doi.org/10.1093/nar/gks940.
[27] F. Chizzolini, M. Forlin, D. Cecchi, S.S. Mansy, Gene position more strongly influences cell-free protein expression from operons than T7 transcriptional promoter strength, ACS Synth. Biol. 3 (2014) 363–371, https://doi.org/10.1021/sb4000977.
[28] S. Zhou, R. Ding, J. Chen, G. Du, H. Li, J. Zhou, Obtaining a panel of cascade promoter-5′-UTR complexes in Escherichia coli, ACS Synth. Biol. 6 (2017) 1065–1075, https://doi.org/10.1021/acssynbio.7b00006.
[29] R.A.J. Darby, A.V. Hine, LacI-mediated sequence-specific affinity purification of plasmid DNA for therapeutic applications, Faseb. J. 19 (2005) 1–20, https://doi.org/10.1096/fj.04-2812fje.
[30] H.M. Salis, E.A. Mirsky, C.A. Voigt, Automated design of synthetic ribosome binding sites to control protein expression, Nat. Biotechnol. 27 (2009) 946–950, https://doi.org/10.1038/nbt.1568.
[31] H.M. Salis, The ribosome binding site calculator, Methods Enzymol. 498 (2011) 19–42, https://doi.org/10.1016/B978-0-12-385120-8.00002-4.
[32] D. Na, S. Lee, D. Lee, Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes, BMC Syst. Biol. 4 (2010), https://doi.org/10.1186/1752-0509-4-71.
[33] D. Na, D. Lee, RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression, Bioinformatics 26 (2010) 2633–2634, https://doi.org/10.1093/bioinformatics/btq458.
[34] C. Gustafsson, S. Govindarajan, J. Minshull, Codon bias and heterologous protein expression, Trends Biotechnol. 22 (2004) 346–353, https://doi.org/10.1016/j.tibtech.2004.04.006.
[35] P. Puigbò, E. Guzmán, A. Romeu, S. Garcia-Vallvé, OPTIMIZER: a web server for optimizing the codon usage of DNA sequences, Nucleic Acids Res. 35 (2007) 126–131, https://doi.org/10.1093/nar/gkm219.
[36] H.G. Menzella, Comparison of two codon optimization strategies to enhance recombinant protein production in Escherichia coli, Microb. Cell Factories 10 (2011) 11–15, https://doi.org/10.1186/1475-2859-10-15.
[37] O. Schlesinger, Y. Chemla, M. Heltberg, E. Ozer, R. Marshall, V. Noireaux, M. H. Jensen, L. Alfonta, Tuning of recombinant protein expression in Escherichia coli by manipulating transcription, translation initiation rates, and incorporation of noncanonical amino acids, ACS Synth. Biol. 6 (2017) 1076–1085, https://doi.org/10.1021/acssynbio.7b00019.
[38] K. Mirzadeh, V. Martínez, S. Toddo, S. Guntur, M.J. Herrgård, A. Elofsson, M.H. H. Nørholm, D.O. Daley, Enhanced protein production in Escherichia coli by optimization of cloning scars at the vector-coding sequence junction, ACS Synth. Biol. 4 (2015) 959–965, https://doi.org/10.1021/acssynbio.5b00033.
[39] L. Berg, R. Lale, I. Bakke, N. Burroughs, S. Valla, The expression of recombinant genes in Escherichia coli can be strongly stimulated at the transcript production level by mutating the DNA-region corresponding to the 5′-untranslated part of mRNA, Microb. Biotechnol. 2 (2009) 379–389, https://doi.org/10.1111/j.1751-7915.2009.00107.x.
[40] S.W. Seo, J.S. Yang, I. Kim, J. Yang, B.E. Min, S. Kim, G.Y. Jung, Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency, Metab. Eng. 15 (2013) 67–74, https://doi.org/10.1016/j.ymben.2012.10.006.
[41] S. Kosuri, D.B. Goodman, G. Cambray, V.K. Mutalik, Y. Gao, A.P. Arkin, D. Endy, G. M. Church, Composability of regulatory sequences controlling transcription and translation in Escherichia coli, Proc. Natl. Acad. Sci. U.S.A. 110 (2013) 14024–14029, https://doi.org/10.1073/pnas.1301301110.
[42] V.K. Mutalik, J.C. Guimaraes, G. Cambray, C. Lam, M.J. Christoffersen, Q.A. Mai, A.B. Tran, M. Paull, J.D. Keasling, A.P. Arkin, D. Endy, Precise and reliable gene expression via standard transcription and translation initiation elements, Nat. Methods 10 (2013) 354–360, https://doi.org/10.1038/nmeth.2404.
[43] V.K. Mutalik, J.C. Guimaraes, G. Cambray, Q.A. Mai, M.J. Christoffersen, L. Martin, A. Yu, C. Lam, C. Rodriguez, G. Bennett, J.D. Keasling, D. Endy, A.P. Arkin,

Quantitative estimation of activity and quality for collections of functional genetic elements, Nat. Methods 10 (2013) 347–353, https://doi.org/10.1038/nmeth.2403.
[44] G. Røkke, E. Korvald, J. Pahr, O. Øyås, R. Lale, BioBrick assembly standards and techniques and associated software tools, Methods Mol. Biol. 1116 (2014) 1–24, https://doi.org/10.1007/978-1-62703-764-8_1.
[45] S.P.O. Santos, A.B.P. Lisboa, F.S.R. Silva, S. Tiwari, V. Azevedo, Á.A. Cruz, E. S. Silva, C.S. Pinheiro, N.M. Alcantara-Neves, L.G.C. Pacheco, Rationally designed hypoallergenic mutant variants of the house dust mite allergen Der p 21, Biochim. Biophys. Acta Gen. Subj. 1866 (2022), https://doi.org/10.1016/j.bbagen.2022.130096.
[46] L.J. Kahl, D. Endy, A survey of enabling technologies in synthetic biology, J. Biol. Eng. 7 (2013) 1, https://doi.org/10.1186/1754-1611-7-13.
[47] R. Kwok, Five hard truths for synthetic biology, Nature 463 (2010) 288–290, https://doi.org/10.1038/463288a.
[48] S.I. Tan, C.C. Hsiang, I.S. Ng, Tailoring genetic elements of the plasmid-driven T7 system for stable and robust one-step cloning and protein expression in broad Escherichia coli, ACS Synth. Biol. 10 (2021) 2753–2762, https://doi.org/10.1021/acssynbio.1c00361.
[49] A. Correa, P. Oppezzo, Tuning different expression parameters to achieve soluble recombinant proteins in E. coli: advantages of high-throughput screening, Biotechnol. J. 6 (2011) 715–730, https://doi.org/10.1002/biot.201100025.
[50] M. Dragosits, D. Nicklas, I. Tagkopoulos, A synthetic biology approach to self-regulatory recombinant protein production in Escherichia coli, J. Biol. Eng. 6 (2012) 2, https://doi.org/10.1186/1754-1611-6-2.
[51] Y. Feng, Q. Xu, T. Yang, E. Sun, J. Li, D. Shi, D. Wu, A novel self-cleavage system for production of soluble recombinant protein in Escherichia coli, Protein Expr. Purif. 99C (2014) 64–69, https://doi.org/10.1016/j.pep.2014.04.001.
[52] C. Mee, M.R. Banki, D.W. Wood, Towards the elimination of chromatography in protein purification: expressing proteins engineered to purify themselves, Chem. Eng. J. 135 (2008) 56–62, https://doi.org/10.1016/j.cej.2007.04.021.
[53] J. Hon, M. Marusiak, T. Martinek, A. Kunka, J. Zendulka, D. Bednar, J. Damborsky, SoluProt: prediction of soluble protein expression in Escherichia coli, Bioinformatics 37 (2021) 23–28, https://doi.org/10.1093/bioinformatics/btaa1102.
[54] B. Pacheco, L. Crombet, P. Loppnau, D. Cossar, A screening strategy for heterologous protein expression in Escherichia coli with the highest return of investment, Protein Expr, Purif 81 (2012) 33–41, https://doi.org/10.1016/j.pep.2011.08.030.

BBA - General Subjects 1866 (2022) 130096

Contents lists available at ScienceDirect

BBA - General Subjects

journal homepage: www.elsevier.com/locate/bbagen

Rationally designed hypoallergenic mutant variants of the house dust mite allergen Der p 21

Sara P.O. Santos [a,b,1], Ayrton B.P. Lisboa [a,c,1], Filipe S.R. Silva [a,b], Sandeep Tiwari [c], Vasco Azevedo [c], Álvaro A. Cruz [d], Eduardo S. Silva [a,e], Carina S. Pinheiro [a,b,e], Neuza M. Alcantara-Neves [a,b,e], Luis G.C. Pacheco [a,b,e,*]

[a] Institute of Health Sciences, Federal University of Bahia, Salvador, BA, Brazil
[b] Post-Graduate Program in Biotechnology, Federal University of Bahia, Salvador, BA, Brazil
[c] Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil
[d] ProAR Foundation and Federal University of Bahia, Salvador, Brazil
[e] Laboratory of Allergology and Acarology (LAA), Federal University of Bahia, Salvador, BA, Brazil

ARTICLE INFO

Keywords:
Allergen-specific immunotherapy
Dermatophagoides pteronyssinus
Der p 21
Recombinant hypoallergen
Allergen engineering
Immunoinformatics

ABSTRACT

Background: Allergic diseases figure among the most common immune-mediated diseases worldwide, affecting more than 25% of the world's population. Allergic reactions can be triggered by house dust mite (HDM) allergens, of which the so-called group 21 of allergens is considered as clinically relevant.
Methods: Herein, we used a structural bioinformatics and immunoinformatics approach to design hypoallergenic mutant variants of the Der p 21 allergen of *Dermatophagoides pteronyssinus*, which were then recombinantly expressed in bacteria and tested for their IgE-reactivities. For this, we scanned the wild-type Der p 21 protein for all possible single amino acid substitutions in key IgE-binding regions that could render destabilization of the major epitope regions.
Results: Four main substitutions (D82P, K110G, E77G, and E87S) were selected to build mutant variants of the Der p 21 allergen, which were produced in their recombinant forms; two of these variants showed reduced reactivity with IgE. Molecular dynamic simulations and immune simulations demonstrated the overall effects of these mutations on the structural stability of the Der p 21 allergen and on the profile of immune response induced through immunotherapy.
Conclusions: When produced in their recombinant forms, two of the Der p 21 mutant variants, namely proteins K110G and E87S, showed significantly reduced IgE reactivities against sera from HDM-allergic individuals ($n = 20$; $p < 0.001$).
General significance: This study successfully translated a rational *in silico* mutagenesis design into low IgE-binding mutant variants of the allergen rDer p 21. These novel hypoallergens are promising to compose next-generation allergen-immunotherapy formulations in near future.

1. Introduction

The most common manifestations of allergic diseases are immunoglobulin E (IgE)-mediated Type I hypersensitivity reactions [1,2], which affect a significant proportion of the global population [3]. One of the main risk factors to the development of Type I hypersensitivity is the sensitization to allergens from the house dust mites (HDM) [4–7]. *Dermatophagoides pteronyssinus and Dermatophagoides farinae* are the major sources of HDM allergens and they are widespread in many regions of the globe, triggering allergic symptoms [6,8,9].

Among the *Dermatophagoides pteronyssinus* allergens, Der p 21 was characterized as a new important allergen, with a high capacity for binding to IgE antibodies and high allergenic activity [10,11]. Pulsawat et al. 2014 suggested that Der p 21 can stimulate the production of IL-8 in airway epithelial cells through TLR2-dependent signaling, similarly to the mechanism described for Der p 2 (a major allergen of

* Corresponding author at: Department of Biotechnology, Institute of Health Sciences, Federal University of Bahia, Salvador, BA 40.110-100, Brazil.
E-mail address: luis.pacheco@ufba.br (L.G.C. Pacheco).
[1] These authors contributed equally to this study.

https://doi.org/10.1016/j.bbagen.2022.130096
Received 26 August 2021; Received in revised form 22 December 2021; Accepted 17 January 2022
Available online 22 January 2022

D. pteronyssinus) [11]. The prevalence of reactivity to Der p 21 has also been investigated by different studies: 25% of subjects in a cohort of HDM-allergic patients in Thailand showed IgE reactivity to this allergen [11]; in Austria, the reported reactivity was 26% [10]; besides, similar percentages of IgE reactivity (*ca.* 20%) were also reported in studies from Singapore [12] and Germany [13]. The Der p 21 sequence has similarities with those of group 5 allergens, but does not show significantly high cross-reactivity. The sequence alignments have shown that Der p 21 and Blo t 5, from the mite *Blomia tropicalis*, have an amino acid (AA) sequence identity of *ca.* 40% [14], while Der p 21 and Der p 5 have *ca.* 31% identity [15]. Sequence studies also revealed that Der p 21 and Blo t 21 share *ca.* 41% of AA sequence identity [15]. These features indicate that Der p 21 is a potential candidate not only for IgE-mediated allergy diagnosis, but also for allergen-specific immunotherapy (AIT) against allergy triggered by *D. pteronyssinus*.

AIT is a traditional treatment for patients suffering from IgE-associated allergy [16] and can be more effective than pharmacotherapy and immunoreactive biologicals in reducing allergy symptoms. The AIT approach takes advantage of the patient's own immune system, and it is the only treatment that shows disease-modifying effects and prevents the progression of severe symptoms [1,3,17]. However, traditional AIT based on crude mite extracts has important limitations due to the potential of inducing severe and systemic side effects, which require up-dose schedules and multiple administrations [3,18,19].

To circumvent these drawbacks, advancements in molecular allergen characterization led to the development of new forms of AIT based on recombinant purified proteins and hypoallergenic peptide derivatives [18,20]. In the hypoallergenization process, allergen derivatives can be produced by identifying the amino acid sequence determinants that are essential to IgE reactivity, and these regions are mutated or excluded at the genetic level, before recombinant protein production. Structural bioinformatics and immunoinformatics tools are generally employed to seek and rationally modify these protein regions in order to maintain the IgG reactivity, while decreasing IgE binding [3,21].

In this study, hypoallergenic recombinant molecules derived from the wild-type (wt) Der p 21 allergen were designed by using a dedicated structural bioinformatics analysis pipeline. Our approach involved *in silico* mutagenesis of the target allergen, allowing for the rational design of hypoallergenic variants. Prior knowledge of the allergen's three-dimensional structure and the location of B cell epitopes are required, enabling a targeted modification of these epitopes, but maintaining the immunogenicity of the allergen. We found that two mutant variants of Der p 21, namely K110G and E87S, presented significant decrease in IgE reactivities when tested with sera of HDM-allergic subjects, demonstrating their potential to compose hypoallergenic AIT formulations.

2. Materials and methods

2.1. Structural bioinformatics and in silico mutagenesis of the Der p 21 allergen

The structural bioinformatics pipeline and the *in silico* design of the of Der p 21 hypoallergenic mutant variants are summarized in the Supplementary Material Fig. S1 and described in the subsequent items.

2.1.1. Sequence retrieval and analyses of protein domains and post-translational modifications

Database entries containing information about Der p 21 were searched in the Allergome relational database (http://www.allergome.org/) [22]. Allergome entries contain links to the sequences in the UniProtKB protein database (http://www.uniprot.org/) [23]. Domain analyses were performed with the Hmmer Profile Search (http://hmmer.org/) against the pfam database (http://pfam.xfam.org/) [24]. Domain prediction was used as a means to try to get further insights on the epitope positions within the Der p 21 allergen, *i.e.* whether or not major epitopes co-localize with predicted protein domains.

2.1.2. Protein tertiary structure prediction, refinement, and validation

Der p 21 wild-type sequences were submitted to the Robetta protein tertiary structure prediction server, using standard parameters (*de novo* modeling and homology modeling) (http://robetta.bakerlab.org/) [25]. Refinement was carried out with FoldIt Standalone [26], which uses a RosettaCM algorithm to refine the protein sidechains and backbone based on free energy score (kcal/mol). Protein tertiary structure validation was then performed using the QMEANDisCo server (https://swissmodel.expasy.org/qmean/) [27] (Supplementary Table S1). 3D-structures visualization was performed with FoldIt and PyMOL *Molecular Graphics System* (https://pymol.org/).

2.1.3. Prediction of B-cell epitopes

Epitopes recognized by B cells were predicted on multiple prediction tools in order to build a consensus and a robust result. CBTope (http://osddlinux.osdd.net/raghava/cbtope/submit.php) [28] analysis were performed with a cut-off value for the SVM Threshold of −0,3 (default), because at this value, sensitivity and specificity were found equal during the development of the software. ElliPro (http://tools.iedb.org/ellipro/) [29] analyses were performed selecting the minimum score of 0.6 (default is 0.5) and maximum distance of 6 Angstroms (default). Both parameters guarantee good sensitivity and, particularly, good specificity. BCEPRED (http://crdd.osdd.net/raghava/bcepred/) [30] analyses were performed using parameters of Antigenic Propensity by Kolaskar et al., 1990 [31]. DiscoTope (https://services.healthtech.dtu.dk/service.php?DiscoTope-2.0) [32] analyses were performed with the threshold for epitope prediction being set at −1.0. At this value, sensitivity was described as 0.30 and specificity as 0.85. Chimera software (https://www.cgl.ucsf.edu/chimera/) was used for the visualization of epitopes.

2.1.4. Predictions of T-cell epitope cleavage sites

MHC-II processed ligands prediction analyses were performed on the IEDB Tools MHCII-NP (http://tools.iedb.org/mhciinp/) [33] server using sequences from Der p 21 wild-type and Der p 21 with all substitutions selected together. Comparison between Der p 21 and the mutants with 4 substitutions was performed to analyze the difference between the MHC II predicted peptides between the variants.

2.1.5. In silico mutagenesis of Der p 21

Amino acid residues with the highest scores on the B-cell epitopes prediction servers results were selected to build a consensus and undergo *in silico* mutagenesis. The positions 77 (glutamic acid), 82 (aspartic acid), 87 (glutamic acid), and 110 (lysine) of the Der p 21 wt allergen were selected to a following a scan for the most destabilizing mutations, performed with the MAESTROweb server (https://biwww.che.sbg.ac.at/maestro/web/) [34]. A manual curation was performed to make sure that mutagenesis would not occur in sites that could lead to a complete disruption in the secondary structure or, consequently, in the tertiary structure of the original protein, such as, for example, proline in an alpha helix. The results of the *in silico* mutagenesis performed with MAESTROweb were then validated using the FoldIt Standalone program. Also, with the FoldIt *in silico* mutagenesis was performed using the '*mutate*' tool. After each substitution, a round of protein structure stabilization was carried in order to submit the protein structure to the effects of the aminoacidic change.

Finally, combinations were made between the most destabilizing mutations identified (E77G, D82P, E87S and K110G) for the production of Der p 21 wt variants: P1.D82P (Asp82Pro substitution), P2.K110G (Asp82Pro + Lys110Gly substitution), P3.E77G (Asp82Pro + Lys110Gly + Glu77Gly substitution) and P4.E87S (Asp82Pro + Lys110Gly + Glu77Gly + Glu87Ser substitution) (Table 1).

2.1.6. Molecular dynamics simulation

To evaluate the structural stabilities of the Der p 21 mutant variants in comparison to the wild-type allergen, we performed a molecular

Table 1
Predictions of the most destabilizing amino acid (AA) substitutions in the Der p 21 allergen and characteristics of the mutant derivatives constructed in this study.

AA position[a]	Original AA	MAESTRO predictions[b]			FoldIt validation[b]	Mutant derivatives (AA substitutions)
		Suggested substitution	$\Delta\Delta G$ (kcal/mol)	*Cpred*[c]	$\Delta\Delta G$ (kcal/mol)	
82	Asp (D)	Pro (P)	1.275	0.904	11.641	P1.D82P (Asp82Pro)
110	Lys (K)	Gly (G)[d]	0.881	0.912	5.785	P2.K110G (Asp82Pro + Lys110Gly)
77	Glu (E)	Gly (G)	2.563	0.815	3.472	P3.E77G (Asp82Pro + Lys110Gly + Glu77Gly)
87	Glu (E)	Ser (S)	2.036	0.844	2.910	P4.E87S (Asp82Pro + Lys110Gly + Glu77Gly + Glu87Ser)

[a] Relative to Der p 21 wild-type sequences retrieved from Allergome DB.
[b] Starting Gibbs free energy value = −176.553 kcal/mol.
[c] Confidence estimation of the calculated $\Delta\Delta G$. The values vary between 0.0 and 1.0, where 1.0 corresponds to a perfect prediction accuracy.
[d] Firstly predicted as Pro (P) but discarded due to potential alpha-helix break.

dynamics simulation experiment. Simulations were carried out by using the GROMACS 2020.3 package [35]. CHARMM36 force field [36] was chosen to perform analysis. Proteins were solvated in a dodecahedron box with TIP3P water molecules. Na^+ and Cl^- ions were added to the system to achieve a neutral charged system. Electrostatic interactions were calculated using the Particle Mesh Ewald for long-range electrostatics method [37]. V-rescale temperature coupling (modified Berendsen) was used to control temperature (310 Kelvin). Berendsen pressure coupling [38] was used to control pressure (1.0 bar). The systems were subjected to unrestrained molecular simulation for 100 ns. Comparative analyses of structural deviations in wild-type and mutant structures were performed. RMSD and RMSF analyses were carried out using GROMACS tools. Graphs were plotted using MATLAB 2020 [39].

2.1.7. Immune simulation analysis

To predict the immune response induced by Derp21 (wt) or by the mutant proteins, we conducted an immune stimulation using C-ImmSim server (http://www.cbs.dtu.dk/services/C-ImmSim-10.1/) [40]. For simulation, two injections containing 1000 vaccine proteins each were given at intervals of one week. Time steps are set at 1 and 21 (each time step is 8 h in real life, and time step 1 is the time of injection =0). The simulation steps have been modified to 1050, and the other parameters were kept as default.

2.2. Design and construction of recombinant plasmids

Table 2 describes all synthetic gene fragments and plasmids constructed in this study. Synthetic genes encoding rDer p 21 wt and the four mutant variants were obtained from Eurofins (Luxemburg), and then cloned using the pCR-Blunt II-TOPO vector (Invitrogen, Carlsbad, CA, EUA). The resulting plasmids were transformed into *E. coli* XL1Blue stain and recombinant clones were cultivated at 37 °C, in Luria-Bertani (LB) broth, with kanamycin. Plasmids were recovered using the Nucleospin plasmid QuickPure kit (Macherey Nagel Co., Düren, Germany), the gene fragments were digested with *SapI* and then subcloned in the expression vector pOPT 2.0 (Table 2) (Supplementary Fig. S2), using standard protocols. Confirmation was performed by digestion with the restriction enzymes *EcoRI* and *PstI* (Promega, Madison, WI, USA) and analysis in a 1% agarose gel.

2.3. Expression, purification, and analysis of recombinant Der p 21 and mutant variants

The expression vectors constructed in this study (Table 2) were transformed into chemically competent *E. coli* BL21(DE3) strains. The cells were grown at 37 °C overnight in LB-agar plates supplemented with 34 μg.mL^{-1} of chloramphenicol. Confirmed clones were incubated in 10 mL LB broth (34 μg.mL^{-1} Cm) and induction of the target recombinant proteins was performed with 1 mM IPTG (Invitrogen, CA, California, EUA), at an $O.D._{600nm} = 0.6$. Cells were then incubated at 37 °C for up to

Table 2
List of synthetic gene fragments, plasmids and bacterial strains used in this study.

	Characteristics	Source
Synthetic gene fragments		
*derp21*wt	387 bp; optimized synthetic gene fragment coding for the recombinant Der p 21 allergen.	This study
mut1	387 bp; optimized synthetic gene fragment coding for the Der p 21 mutant variant P1.D82P.	This study
mut2	387 bp; optimized synthetic gene fragment coding for the Der p 21 mutant variant P2.K110G.	This study
mut3	387 bp; optimized synthetic gene fragment coding for the Der p 21 mutant variant P3.E77G.	This study
mut4	387 bp; optimized synthetic gene fragment coding for the Der p 21 mutant variant P4.E87S.	This study
Strains		
E. coli XL1Blue	*endA1 gyrA96(nalR) thi-1 recA1 relA1 lac glnV44 F′[:: Tn10 proAB+ lacIq Δ(lacZ)M15] hsdR17(rK- mK+)*	Stratagene
E. coli BL21 (DE3)	B F– ompT gal dcm lon hsdSB(rB–mB–) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS)	Stratagene
Plasmids		
pCR-Blunt II-TOPO	ColE1/ Kan^R, Neo^R; *lacZalpha*, *ccdB*	Invitrogen
pTP1: *derp21*wt	ColE1/ Kan^R, Neo^R; fragment encoding the recombinant Der p 21 allergen (rDer p 21), flanked by *SapI* sites.	This study
pTP2:*mut1*	ColE1/ Kan^R, Neo^R; cloned fragment encoding the Der p 21 mutant variant P1.D82P, flanked by *SapI* sites.	This study
pTP3:*mut2*	ColE1/ Kan^R, Neo^R; cloned fragment encoding the Der p 21 mutant variant P2.K110G, flanked by *SapI* sites.	This study
pTP4:*mut3*	ColE1/ Kan^R, Neo^R; cloned fragment encoding the Der p 21 mutant variant P3.E77G, flanked by *SapI* sites.	This study
pTP5:*mut4*	ColE1/ Kan^R, Neo^R; cloned fragment encoding the Der p 21 mutant variant P4.E87S, flanked by *SapI* sites.	This study
pOPT2.0	pUC ori/ Cm^R; Expression vector containing an optimized T7 promoter, strong RBS, fragment encoding GFP, 6× His-tag, T7 term. SapI-mediated scarless cloning.	This study
pOPT: *derp21*wt	pOPT2.0 containing a cloned gene fragment encoding the recombinant Der p 21 allergen (rDer p 21).	This study
pOPT:*mut1*	pOPT2.0 containing a cloned gene fragment encoding the Der p 21 mutant variant P1.D82P.	This study
pOPT:*mut2*	pOPT2.0 containing a cloned gene fragment encoding the Der p 21 mutant variant P2.K110G.	This study
pOPT:*mut3*	pOPT2.0 containing a cloned gene fragment encoding the Der p 21 mutant variant P3.E77G.	This study
pOPT:*mut4*	pOPT2.0 containing a cloned gene fragment encoding the Der p 21 mutant variant P4.E87S.	This study

20 h. For recombinant protein purification, centrifuged bacterial pellets were collected at different time points during the induction protocol. Pellets were resuspended with 50 mM phosphate buffer, with or without 6 M urea (pH 8.0). Resuspended pellets were sonicated thrice, for 30 s at 50 Hz. His-tagged recombinant proteins were purified using the MagneHis™ Protein Purification system (Promega, Madison, WI, EUA), following manufacturer's instructions. Recombinant proteins were resolved in 12% SDS-PAGE and confirmed by Western blot using mouse anti-His IgG (GE Healthcare, 1:4000). Protein concentration was measured with the Qubit™ Protein Assay kit (Thermo Fisher Scientific, Waltham, MA, EUA).

2.4. Sera from HDM-allergic patients

In this study, two sets of sera were used. The first group of donors was composed of sensitized individuals, from both genders, aged ≥18 years, and living in the city of Salvador, Brazil. The second set of sera was obtained from individuals enrolled in Asthma and Allergic Rhinitis Control Program in Bahia (ProAR) [41]. The allergy diagnosis, in both cases, was based on allergy symptoms (rhinitis, conjunctivitis, and/or asthma), positive skin prick test to *Dermatophagoides pteronyssinus*, and serum levels of IgE anti-*D. pteronyssinus* using immunoCAP (> 0,7 kUA/L). Sera from nonallergic individuals were used as control. The study was approved by the Ethics Committee on Research of the Faculty of Medicine of the Federal University of Bahia (CAAE 45376814.0.0000.5577). Blood samples were taken after informed consent was obtained from patients.

2.5. IgE reactivities of recombinant Der p 21 wt and of the mutant variants

Firstly, a *Dot blot* assay was performed to analyze the IgE reactivity to the Der p 21 wt and the mutant variants. For this, 2.5 μg of each recombinant protein was applied to a nitrocellulose membrane (GE Healthcare, Waukesha, WI, USA). Likewise, a crude extract of *Dermatophagoides pteronyssinus* and a purified recombinant Der p 2 protein were used as positive controls. After 1.5 h incubation with the antigens, membranes were blocked for 2 h at room temperature using a 3% skimmed milk and 5% tween-20 PBS solution. Then, membranes were incubated with pools of sera (1:5 diluted) at 4 °C, for 16 h. A sera pool of five samples from sensitized patients was used, and another pool of five non-allergic patients was used as control. Following sera incubation, the membranes were incubated with a biotinylated anti-human IgE antibody (1:500, BD Biosciences, San Jose, CA) for 1 h at room temperature and then incubated with Streptavidin-HRP (1:2500, BD Biosciences, San Jose, CA) for 45 min. The reacted bands were visualized using Enhanced Chemiluminescence (ECL) (GE Healthcare, Waukesha, WI, USA) and ImageQuant LAS 4000 (GE Healthcare, Waukesha, WI, USA). ImageJ software was also used to analyze the reacted spots.

The IgE reactivities of the different recombinant proteins were also evaluated by ELISA assays, according to a protocol previously developed by our group [4]. Briefly, high binding plates (Greiner Bio-One MICROLON™) were coated with 5.0 $\mu g.mL^{-1}$ of recombinant Der p 21 wt or the mutant variants, in a carbonate-bicarbonate buffer (pH 9.6–9.8), for 12 h at 4 °C. The plates were also blocked using a 10% BSA and 0.05% tween-20 PBS solution. The plate wells containing each Der p 21 variant was incubated with the positive or negative serum pool (diluted 1:5 in the blocking solution) for 16 h at 4 °C. The plates were then incubated with biotinylated anti-human IgE antibody (1:2000, BD Biosciences, San Jose, CA) for 1 h at room temperature, followed by Streptavidin-HRP (1:1000, BD Biosciences, San Jose, CA) for 1 h at room temperature. Finally, 0.1 mg.mL-1 TMB and 30% hydrogen peroxide were added to the plates for the reaction development. The optical density at 450 nm was measured by using a plate reader (Multiskan™ FC Microplate Photometer, Thermo Fisher Scientific, Vantaa, Finland).

For these ELISA assays, firstly, sera from 30 sensitized donors were used to access their individual IgE reactivities against a crude extract of *Dermatophagoides pteronyssinus* or rDer p 21. Following the confirmation of the donors with the highest reactivities to rDer p 21, the sera were again used to evaluate the IgE reactivities of the recombinant mutant variants.

2.6. Statistical analysis

GraphPad Prism software was used to plot the IgE reactivity data and to perform statistical analysis. Normality was verified with Shapiro–Wilk normality test and groups were analyzed by Wilcoxon test signed rank test. A value of $p < 0.05$ was considered as statistically significant.

3. Results and discussion

Immunotherapy for HDM allergy using hypoallergenic allergens is currently considered as a promising alternative for more effective AIT [42,43]. To achieve this, recent studies have described protein engineering strategies to produce hypoallergenic mutant derivatives with reduced IgE binding, which retain epitopes recognized by T cells and induce blocking IgG antibodies [19–21,42,44,45]. The overall aim of this strategy is to overcome the negative effects of allergen-specific immunotherapy, by maintaining the immunogenicity of AIT without increasing its allergenicity, then improving the risk-benefit ratio [46,47].

In this study, we selected the Der p 21 allergen of *D. pteronyssinus* for hypoallergenization using the *in silico* mutagenesis approach, due to its high frequency of IgE recognition and allergenic activity [10,11]. Previous studies have already shown that *in silico* designed hypoallergens can exhibit a significant reduction in IgE binding, while retaining T-cell stimulating properties [42,46,48,49]. In addition to producing safer and more effective recombinant proteins for allergen immunotherapy, the *in silico* rational design of hypoallergens can also contribute to a reduction in the costs for obtaining the final product, given the reduction in experimental steps if compared to more traditional protein engineering methods such as gene shuffling [34,48,50,51].

Firstly, we performed tridimensional structure modeling, refinement, and validation of the Der p 21 wild-type allergen (Fig. 1a; Supplementary Fig. S1). Following predictions of linear and discontinuous B-cell epitopes (Supplementary Material Table S2 and Fig. S3), four amino acid residues were selected to undergo substitutions aimed at destabilization of the epitope regions of Der p 21, namely Glu(E)77, Asp (D)82, Glu(E)87, and Lys(K)110 (Fig. 1a-b). Three of these amino acid residues are present at the predicted conformational epitope E3 (Supplementary Fig. S4; Supplementary Table S2). Importantly, this region also co-localizes with an experimentally validated major conformational IgE-epitope of the Der p 21 allergen, termed peptide region P4 in the study by Curin and collaborators [44]. The computational screening for destabilizing mutations that affect B-cell epitopes was firstly used with success by Thalhamer and collaborators to generate hypoallergenic variants of the major pollen allergens of birch, Bet v 1, and timothy grass, Phl p 5 [42]. In our study, we have used the MAESTRO software [34] for the initial screening of the most destabilizing amino acid substitutions for each position, and then potential combinations of mutations were also evaluated with the aid of the FoldIt Standalone program [26]. Basically, the MAESTRO tool uses machine learning to predict changes in the Gibbs free energy ($\Delta\Delta G$) of a given protein caused by single or combined amino acid substitutions. Additionally, it is possible to scan for the most stabilizing or destabilizing mutations. Positive values of $\Delta\Delta G$ indicate destabilization and MAESTRO also provides a confidence estimation for $\Delta\Delta G$ predictions [34].

The substitution of an aspartic acid residue at position 82 of the wild-type Der p 21 protein by a proline was selected by the MAESTRO software and then ranked by FoldIt as the most destabilizing mutation (Table 1); other changes that were also ranked as the top destabilizing substitutions were: change of a lysine at position 110 by a glycine;

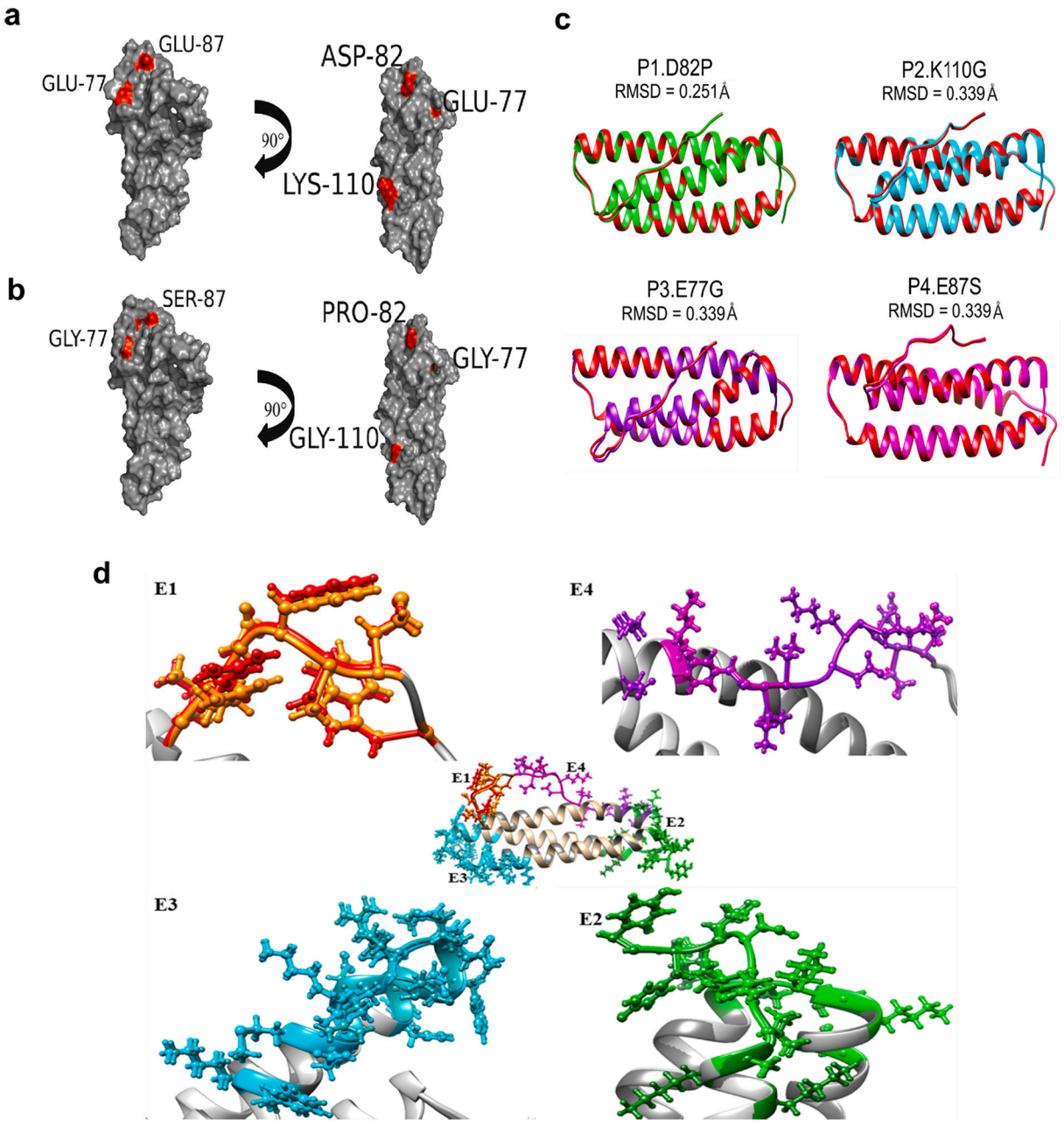


Fig. 1. *In silico* mutagenesis of the Der p 21 allergen. (a) The four amino acid positions selected for mutagenesis. (b) Amino acid substitutions predicted as the most destabilizing in the conformational epitope regions. (c) Superimpositions of the 3D models of the four mutant variants with the Der p 21 wild-type allergen. Root-mean squared deviation (RMSD) values are indicated. (d) The four IgE-epitope regions (E1-E4) identified with high confidence in Der p 21. The zoomed in regions show superimpositions of the mutant variant containing the four substitutions (P4.E87S) with the Der p 21(wt).

change of glutamic acid at position 77 by a glycine; and change of glutamic acid at position 87 by a serine (Table 1). These substitutions were combined to generate the four mutant derivatives of Der p 21: P1.D82P, P2.K110G, P3.E77G, and P4.E87S (Fig. 1a-b; Table 1). Importantly, the selected mutations do not significantly alter the three helical bundle structure of the native Der p 21 allergen (Fig. 1c-d). A similar structure has been experimentally demonstrated for homologous proteins of the groups 21 and 5 of allergens [52]. Interestingly, either drastic or non-drastic structural changes have been linked with the hypoallergenicity of a derivative for AIT [21,45,53–55]. In the case of some hypoallergenic hybrid to treat allergy caused by HDM, there was no drastic changes in comparison with the wild-type allergen [21,55].

In addition to evaluating the effect of the proposed substitutions on the structure of the epitope regions using the folded protein structures, we employed a molecular dynamics simulation (MDS) strategy to study the potential of destabilization of these mutations on the overall structure of the Der p 21 allergen in a dynamic environment. MDS is a tool that has been commonly adopted in computational protein engineering in recent years to predict the effects of newly introduced mutations on the structural stability of engineered proteins [56–58]. It is often employed with the aim of improving the stability of mutant derivatives of biotechnologically important proteins [56,58,59]. In our study, this strategy contributed to the detection of a significant structural instability introduced by substitution of the glutamic acid at position 77 of Der p 21 (Fig. 2). Results in Fig. 2a present dynamic simulations of 100 ns duration for each of the four mutant derivatives, in addition to the wild-type allergen. The backbone root mean squared deviation (RMSD) of the Der p 21 (wt) protein showed a rapid increase in the first 10 ns of

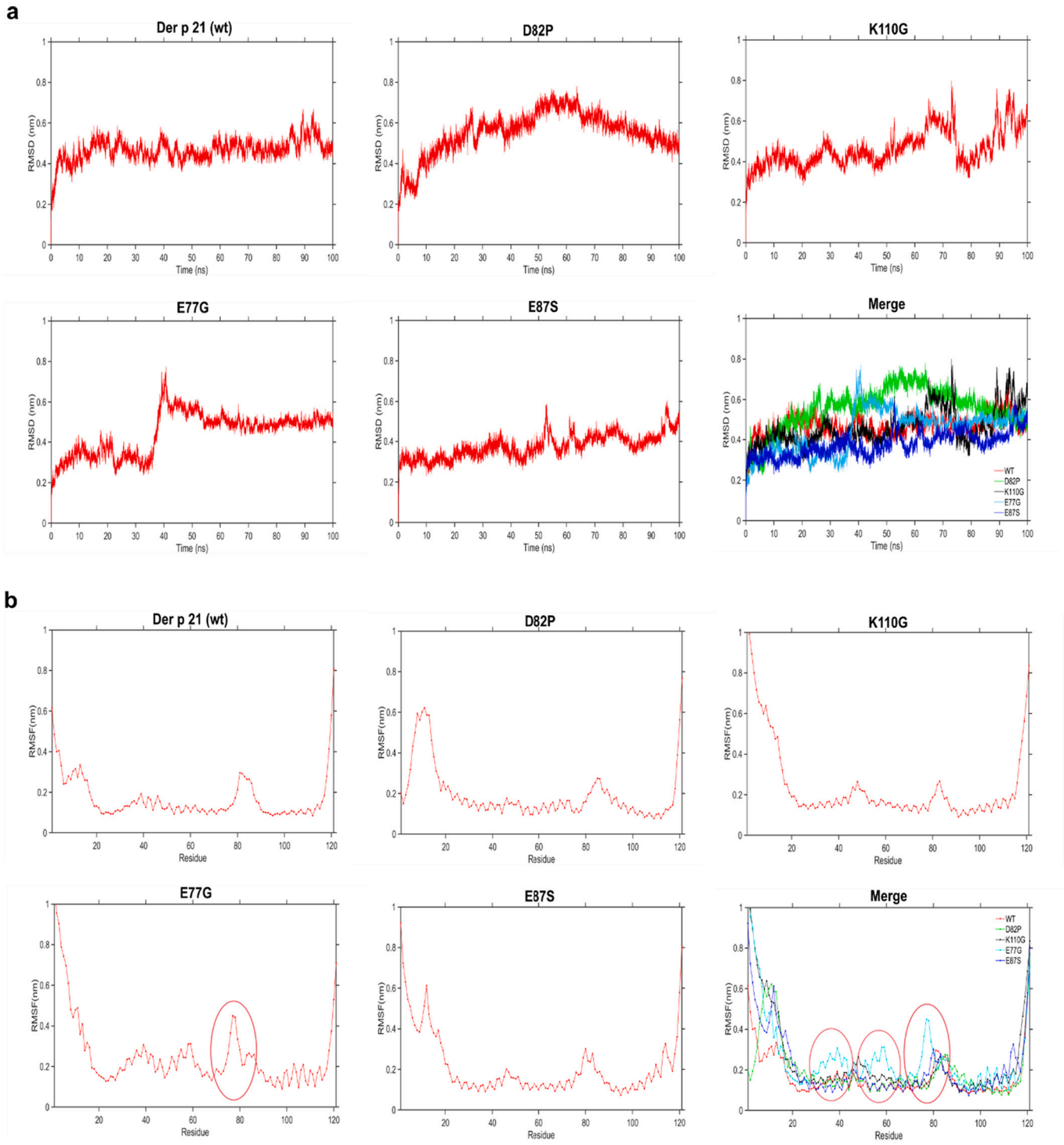


Fig. 2. Molecular dynamics simulation of the Der p 21 allergen and the four mutant variants. (a) RMSD values over a 100 ns simulation. (b) RMSF values per residue.

the simulation and rapidly stabilized at around 4 Å (0.4 nm), with a gradual increase to *ca.* 6 Å over the course of the experiment. The RMSD of the mutant derivatives, in turn, showed large deviations from structural stability over the course of dynamic simulation (Fig. 2a), even though all proteins had been classified as generally stable according to the instability indexes predicted by physicochemical parameters computed with the ProtParam tool (data not shown). It is noteworthy that allergens are generally considered as highly structured proteins, when compared to non-allergenic proteins, and smaller RMSD fluctuations are expected in regions that correlate with major IgE epitopes [60].

Our Root mean square fluctuations (RMSF) per residue results (Fig. 2b) clearly demonstrate that the specific amino acid substitutions performed in this study in the Der p 21 protein were responsible by large fluctuations in the epitope regions, with a significantly higher destabilizing effect attributed to the E77G mutation when introduced in the mutant protein 3 (P3.E77G) (Fig. 2b, merged panel). Interestingly, introduction of the fourth mutation (E87S) exerted a stabilizing effect on the mutant protein (Fig. 2a and b).

As aforementioned, an essential feature of a hypoallergenic protein, in order to be successfully used in next-generation allergy vaccines, is

the retention of a helper T-cell reactivity comparable to that induced by the wild type allergen, despite the reduced IgE binding ability. In our study, a prediction of MHC class II bound T-cell epitopes of the Der p 21 wt allergen and the four mutant derivatives showed a conservation on cleavage sites following the four amino acid substitutions (Supplementary Table S3). Moreover, we used the C-ImmSim server to perform a dynamic immune simulation of the potential response induced by a prime-boost vaccination scheme (days 0 and 7) with the wild-type Der p 21 allergen and the four mutant variants (Fig. 3). Fig. 3a and b shows the predicted profiles of B-cells and T-helper cells induced by vaccination with the wild-type Der p 21 allergen, over a course of nearly a year. There is a predominantly induction of a T helper 1 (Th 1) profile of cells (Fig. 3; inset chart). Notably, a two-vaccinations regimen with the wild-type Der p 21 allergen is expected to induce high levels of INF-gamma, IL-12, TGF-beta, and IL-10 (Fig. 3c). No major alterations are predicted in the productions of cytokines and interleukins, when comparing the profiles induced by vaccination with the wild-type and the four mutant variants (Fig. 3d).

To evaluate whether the Der p 21 mutant variants designed by *in silico* mutagenesis would in fact present reduced IgE-binding activity when incubated with sera of allergic patients, we firstly implemented an effort to optimize the production and purification of the recombinant forms of these proteins expressed in bacteria. Recombinant Der p 21 (rDer p 21) had already been produced and recovered in its soluble form, without fusion to solubility or purification tags, by previous studies using *Escherichia coli* and the yeast *Pichia pastoris* as expression hosts [10,11]. In this study, we chose to introduce a 6×-His purification tag at the C-terminus of the five recombinant proteins, through translational fusion following scarless in-frame cloning at the pOPT2.0 expression vector (Table 2; Fig. 4a). Fig. 4b shows a schematic representation of the expected rDer p 21 protein (15,846 Da) expressed from pOPT2.0, following induction by IPTG. The amino acid substitutions in the recombinant mutant variants are highlighted in the figure, and all have been predicted as soluble by analyses with the tool Protein-Sol [61,62], with solubility indexes higher than 0.6. The solubility of recombinant proteins in physiological buffers is critical for efficient and reliable production of therapeutic biomolecules, in order to overcome a well-known bottleneck in the biotechnology industry [63,64].

In fact, the successful recombinant *E. coli* clones expressed rDer p 21 and three of the mutant variants (P1.D82P, P2.K110G, and P4.E87S) in the soluble and insoluble fractions, as confirmed by 12% SDS-PAGE (Fig. 4c). These four recombinant proteins were purified to completion by affinity chromatography (Fig. 4d) with final protein yields ranging from 55.6 mg / L of culture to 60.0 mg / L of culture (Supplementary Table S4). Notably, the mutant protein 3 (P3.E77G), which was predicted as the most unstable by the dynamics simulation analysis, could not be successfully purified when expressed in *E. coli*, even when key expression parameters were modified, as suggested elsewhere [65,66] (Supplementary Fig. S5). Corroborating previous studies that also produced rDer p 21 [10,11] or rDer f 21 [52], the ortholog protein from

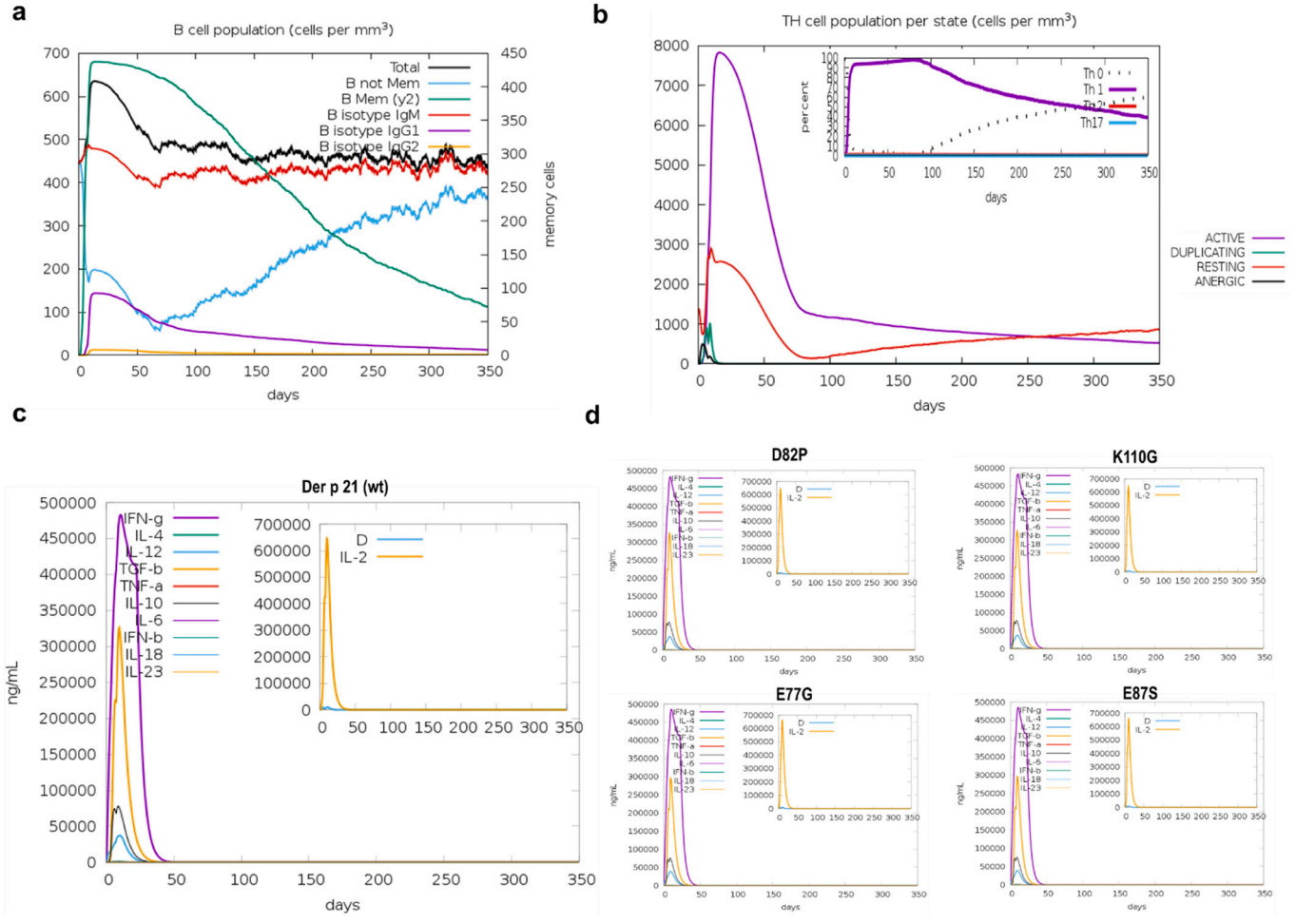


Fig. 3. Simulation of the immune response induced by vaccination with Der p 21 and the four mutant variants. (a) and (b) The profiles of B-cells and T-helper cells induced following a prime-boost vaccination regimen with the wild-type Der p 21 protein at days 0 and 7. The inset chart in (b) indicates the pattern of cellular immune response induced by vaccination. (c) Predicted cytokines and interleukins induced by vaccination with Der p 21(wt) over time. (d) Predicted cytokines and interleukins induced by vaccination with the mutant variants.

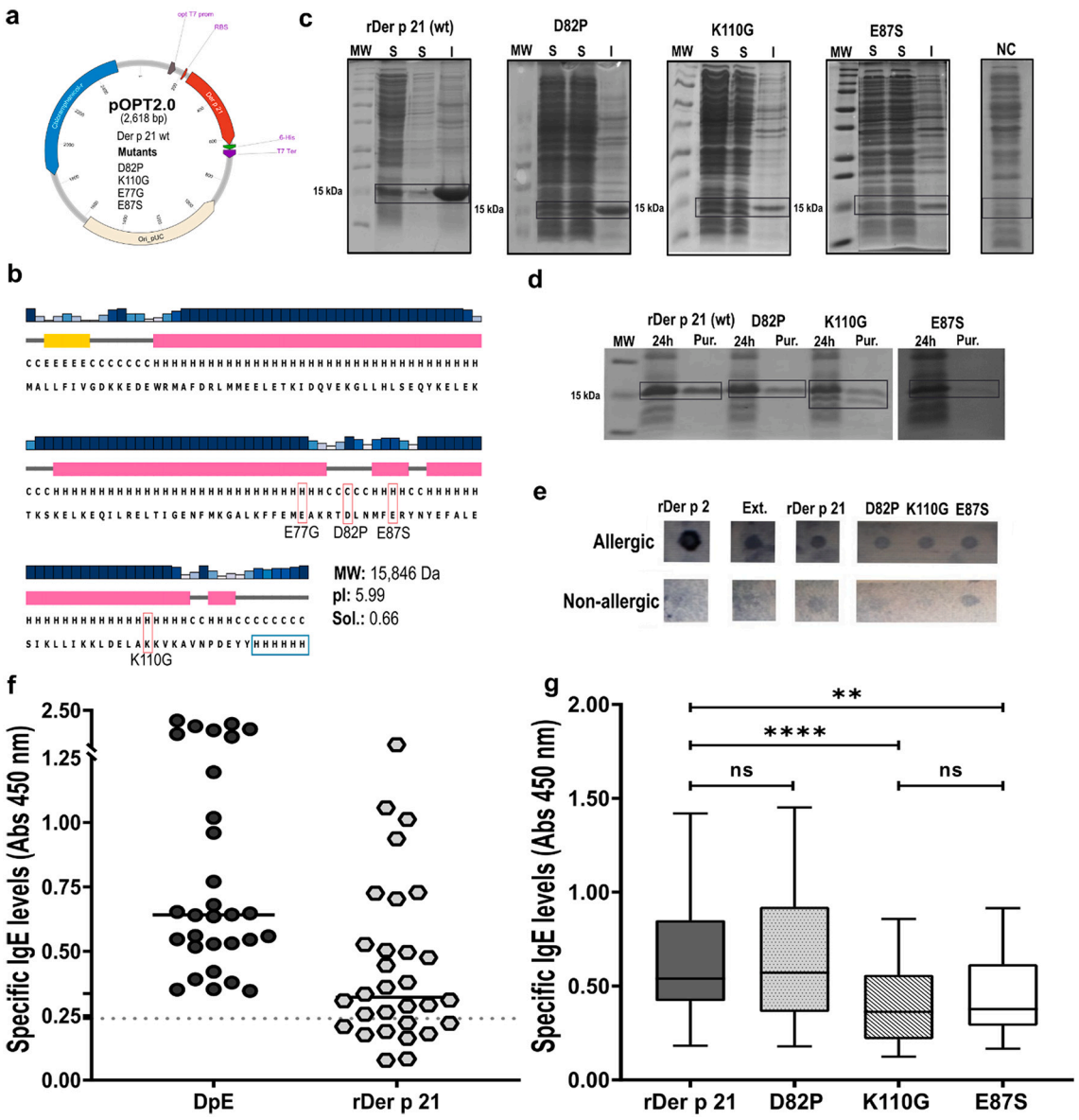


Fig. 4. Recombinant expression of the Der p 21 mutant variants and evaluation of IgE binding ability. (a) Schematic of the cloning of the gene fragments coding for rDer p 21 and the mutant variants in the pOPT2.0 expression vector. (b) Primary and predicted secondary structure of the recombinant proteins expressed in *E. coli*. The included 6×-His tag is indicated at the C-terminus. The pink regions indicate helical structures predicted by PSIPRED server. Darker blue bars indicate the confidence of secondary structure prediction per region. Molecular weight and isoelectric point of the rDer p 21 (wt) were calculated by ProtParam. Solubility score was calculated by Protein-Sol server. (c) 12% SDS-PAGE showing the expression profiles of rDer p 21 and the mutant variants. S = soluble fraction; I = insoluble fraction. NC = negative control, non-transformed BL21(DE3). (d) Purification of the recombinant proteins by affinity chromatography, following 24 h induction. (e) Dot blot analysis with sera from 5 atopic individuals. (f) IgE reactivities of atopic sera ($n = 30$) against mite crude extract (DpE) or rDer p 21. (g) IgE reactivities of rDer p 21 and the mutant variants against highly reactive atopic sera ($n = 20$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Dermatophagoides farinae, our Western blot results using an anti-His monoclonal antibody also suggest that the recombinant proteins may be found in solution partially as dimeric forms (Supplementary Fig. S6). Dimerization and aggregation can sometimes interfere in IgE-binding and hide epitopes in the allergen, leading to possible untrustworthy sensitization rates and usually reduction in reaction [67,68]. However, through dot blot analysis, we demonstrated that the wild-type rDer p 21 and all the three produced mutant variants are recognized by sera from allergic individuals ($n = 5$) (Fig. 4e), but react with different intensities. To quantify the differential levels of IgE binding, we then performed antigen-specific ELISA assays with these proteins, evaluating a set of sera from 30 *Dermatophagoides pteronyssinus*-sensitized patients. As shown in Fig. 4f, all donors reacted to the crude mite extract above the cut-off (0.3362), but only 20 of those reacted to rDer p 21. This IgE reactivity of 66.67% was higher than previous studies in both tropical and temperate climate regions of the globe [10,11] and classifies the allergen as a major allergen in Brazilian sensitized patients. Even though the allergen was partially dimerized in solution and considering the strong reaction in half of the rDer p 21-reactive individuals (Fig. 4f), this finding may lead us to believe this allergen may be included in the allergy diagnosis in the future, especially molecular based-diagnostic tools, an important trend in the field [69,70]. Moreover, although previous studies did find limited cross-reactivity between Der p 21 and Der p 5 [44], studies on this topic for rBlo t 21 and rBlo t 5 cross-reaction with Der p 21 are still warranted. Considering the high inhibition behavior of Der f 21 against Blo t 5 and Blo t 21 IgE binding in another

study [71], cross-reactivity between these two major *Blomia tropicalis* allergens and Der p 21 is a possibility that could directly interfere in allergy diagnosis.

The twenty reactive sera of allergic patients were again evaluated for rDer p 21 IgE reaction, but in comparison with IgE reaction against the mutant variants D82P, K110G, and E87S (Fig. 4g). There is a significant reduction in the IgE reactivities of the mutant proteins K110G and E87S, whereas the variant D82P, containing a single amino acid substitution, does not significantly differ from the wild-type recombinant allergen rDer p 21 (Fig. 4g). The reactivities of the proteins K110G and E87S are not significantly different, highlighting that they are similarly hypoallergenic. The low IgE binding is a key feature in the production of an effective and safe hypoallergenic protein, suitable to be employed in AIT [19,42,72–77]. The fact that the proteins containing combinations of amino acid substitutions (K110G and E87S) presented reduced IgE reactivities when compared to the single substitution in D82P corroborates similar studies with the Blo t 5 allergen of *Blomia tropicalis*, in which it was observed that the reduction in IgE binding was more significant in variants that contained multiple mutations [78]. In addition, our group has recently shown that mutation of four amino acid residues in a hybrid protein led to a hypoallergenic derivative with low IgE reactivity in sera from both allergic non-asthmatic and allergic asthmatic donors, including low avidity features and lack of basophils' activation [21]. Noteworthy, in the present study the inclusion of the fourth mutation E87S contributed to stabilize the overall structure of the recombinant protein, without changing the profiles of low IgE reactivity and the predicted T-cell activation; this might represent an advantage for further exploration of this mutant variant to compose a potential hypoallergenic vaccine formulation for house dust mite allergies.

Although further studies are still warranted, the present study successfully translated a rational *in silico* mutagenesis design into low IgE-binding mutant variants of the allergen rDer p 21. Additionally, we obtained high yields of recombinant expression and demonstrated, for the first time, that Der p 21 can be considered a major allergen in a tropical city of Brazil.

Declaration of Competing Interest

S.P.O.S., A.B.P.L., F.S.R.S., E.S.S., C.S.P., N.M.A.N, and L.G.C.P. have filled a patent at Federal University of Bahia containing partial results of this project. It is currently pending at the National Institute of Industrial Property (INPI) of Brazil. The authors declare no other competing interests.

Acknowledgements

S.P.O.S. was recipient of a PhD scholarship from CAPES foundation, Brazil. V.A., A.A.C., N.M.A.N, and L.G.C.P. are recipients of research fellowships from CNPq-Brazil. This study was partially funded by FAPESB, through research grant #PNE0007/2014.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagen.2022.130096.

References

[1] Y. Zhernov, M. Curin, M. Khaitov, A. Karaulov, R. Valenta, Recombinant allergens for immunotherapy: state of the art, Curr. Opin. Allergy Clin. Immunol. 19 (2019) 402–414.
[2] S. Blank, C. Hilger, Editorial: novel advances in allergy diagnosis and treatment, Front. Immunol. 12 (2021) 10–12.
[3] R. Valenta, R. Campana, V. Niederberger, Recombinant allergy vaccines based on allergen-derived B cell epitopes, Immunol. Lett. 189 (2017) 19–26.
[4] O.P. de Melo-Neto, F.B. Magalhães, J.C.M. Ponte, F.A.B. Felipe, M.C.A. dos Santos, G. dos Santos Lima, Á.A. Cruz, C.S. Pinheiro, L.C. Pontes-de-Carvalho, N. M. Alcantara-Neves, K.A. dos Carvalho, Blomia tropicalis Blo t 5 and Blo t 21 recombinant allergens might confer higher specificity to serodiagnostic assays than whole mite extract, BMC Immunol. 14 (2013).
[5] M.A. Calderón, A. Linneberg, J. Kleine-Tebbe, F.De Blay, D.Hernandez Fernandez De Rojas, J.C. Virchow, P. Demoly, Respiratory allergy caused by house dust mites: What do we really know? J. Allergy Clin. Immunol. 136 (2015) 38–48.
[6] N. Acevedo, J. Zakzuk, L. Caraballo, House dust mite allergy under changing environments, Allergy. Asthma Immunol. Res. 11 (2019) 450.
[7] J.D. Miller, The role of dust mites in allergy, Clin. Rev. Allergy Immunol. 57 (2019) 312–329.
[8] K. Sade, D. Roitman, S. Kivity, Sensitization to dermatophagoides, blomia tropicalis, and other mites in atopic patients, J. Asthma 47 (2010) 849–852.
[9] R. Limão, A. Spínola Santos, L. Araújo, J. Cosme, F. Inácio, E. Tomaz, A. Ferrão, N. Santos, A. Sokolova, A. Môrete, H. Falcão, L. Cunha, A. Ferreira, A. Bras, F. Ribeiro, C. Lozoya, P. Leiria Pinto, S. Prates, J. Plácido, A. Coimbra, et al., Molecular sensitization profile to dermatophagoides pteronyssinus dust mite in Portugal, J. Investig. Allergol. Clin. Immunol. 32 (2020) 1–18.
[10] M. Weghofer, Y. Dall'Antonia, M. Grote, A. Stöcklinger, M. Kneidinger, N. Balic, M. T. Krauth, E. Fernández-Caldas, W.R. Thomas, M.Van Hage, S. Vieths, S. Spitzauer, F. Horak, D.I. Svergun, P.V. Konarev, P. Valent, J. Thalhamer, W. Keller, R. Valenta, S. Vrtala, Characterization of Der p 21, a new important allergen derived from the gut of house dust mites, Allergy Eur. J. Allergy Clin. Immunol. 63 (2008) 758–767.
[11] P. Pulsawat, M. Theeraapisakkun, E. Nony, M. Le Mignon, K. Jain, A. Buaklin, J. Wongpiyabovorn, K. Ruxrungtham, A. Jacquet, Characterization of the house dust mite allergen der p 21 produced in Pichia pastoris, Protein Expr. Purif. 101 (2014) 8–13.
[12] M.I. Kidon, C.W. Chin, L.W. Kang, O.T. Ching, T.Y. Seng, W.K. Ning, A.C. Angus, O. S. Theng, G.Y. Feng, K. Reginald, B.X. Zhi, S.H. Shen, C.F. Tim, Mite component-specific IgE repertoire and phenotypes of allergic disease in childhood: the tropical perspective, Pediatr. Allergy Immunol. 22 (2011) 202–210.
[13] D. Posa, S. Perna, Y. Resch, C. Lupinek, V. Panetta, S. Hofmaier, A. Rohrbach, L. Hatzler, L. Grabenhenrich, O. Tsilochristou, K.W. Chen, C.P. Bauer, U. Hoffman, J. Forster, F. Zepp, A. Schuster, U. Wahn, T. Keil, S. Lau, S. Vrtala, et al., Evolution and predictive value of IgE responses toward a comprehensive panel of house dust mite allergens during the first 2 decades of life, J. Allergy Clin. Immunol. 139 (2017) 541–549, e8.
[14] J.C. Bessot, G. Pauli, Mite allergens: an overview, Eur. Ann. Allergy Clin. Immunol. 43 (2011) 141–156.
[15] Y.F. Gao, D.Y. Wang, T.C. Ong, S.L. Tay, K.H. Yap, F.T. Chew, Identification and characterization of a novel allergen from blomia tropicalis: blo t 21, J. Allergy Clin. Immunol. 120 (2007) 105–112.
[16] Y. Dorofeeva, I. Shilovskiy, I. Tulaeva, M. Focke-Tejkl, S. Flicker, D. Kudlay, M. Khaitov, A. Karsonova, K. Riabova, A. Karaulov, R. Khanferyan, W.F. Pickl, T. Wekerle, R. Valenta, Past, present, and future of allergen immunotherapy vaccines, Allergy EurJ. Allergy Clin. Immunol. 76 (2021) 131–149.
[17] R. Campana, K. Marth, P. Zieglmayer, M. Weber, C. Lupinek, Y. Zhernov, O. Elisyutina, M. Khaitov, E. Rigler, K. Westritschnig, U. Berger, M. Wolkersdorfer, F. Horak, F. Horak, R. Valenta, Vaccination of nonallergic individuals with recombinant hypoallergenic fragments of birch pollen allergen Bet v 1: safety, effects, and mechanisms, J. Allergy Clin. Immunol. 143 (2019) 1258–1261.
[18] E.S. da Silva, C.S. Pinheiro, L.G.C. Pacheco, N.M. Alcantara-Neves, Dermatophagoides spp hypoallergens design: what has been achieved so far? Expert Opin. Ther. Pat. 30 (2020) 163–177.
[19] E.S. da Silva, C.S. Pinheiro, C.M. Quintella, F. Ferreira, L.G. Luis, N.M. Alcântara-Neves, Advances in patent applications related to allergen immunotherapy, Expert Opin. Ther. Pat. 26 (2016) 657–668.
[20] O. Akinfenwa, A. Rodríguez-Domínguez, S. Vrtala, R. Valenta, R. Campana, Novel vaccines for allergen-specific immunotherapy, Curr. Opin. Allergy Clin. Immunol. 21 (2021) 86–99.
[21] E.S. da Silva, L. Aglas, C.S. Pinheiro, E.M.M. de Andrade Belitardo, E.F. Silveira, S. Huber, R.T. Torres, M. Wallner, P. Briza, P. Lackner, J. Laimer, L.G.C. Pacheco, Á.A. Cruz, N.M. Alcântara-Neves, F. Ferreira, A hybrid of two major blomia tropicalis allergens as an allergy vaccine candidate, Clin. Exp. Allergy 50 (2020) 835–847.
[22] A. Mari, C. Rasi, P. Palazzo, E. Scala, Allergen databases: current status and perspectives, Curr. Allergy Asthma Rep. 9 (2009) 376–383.
[23] T.U. Consortium, UniProt: a worldwide hub of protein knowledge, Nucleic Acids Res. 47 (2019) D506–D515.
[24] J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G.A. Salazar, E.L. L. Sonnhammer, S.C.E. Tosatto, L. Paladin, S. Raj, L.J. Richardson, R.D. Finn, A. Bateman, Pfam: the protein families database in 2021, Nucleic Acids Res. 49 (2021) D412–D419.
[25] Y. Song, F. Dimaio, R.Y.R. Wang, D. Kim, C. Miles, T. Brunette, J. Thompson, D. Baker, High-resolution comparative modelingwith RosettaCM, Structure 21 (2013) 1735–1742.
[26] R. Kleffner, J. Flatten, A. Leaver-Fay, D. Baker, J.B. Siegel, F. Khatib, S. Cooper, Foldit standalone: a video game-derived protein structure manipulation interface using Rosetta, Bioinformatics 33 (2017) 2765–2767.
[27] P. Benkert, S.C.E. Tosatto, D. Schomburg, QMEAN: a comprehensive scoring function for model quality assessment, Proteins Struct. Funct. Bioinforma. 71 (2008) 261–277.
[28] H.R. Ansari, G.P. Raghava, Identification of conformational B-cell epitopes in an antigen from its primary sequence, Immunome Res. 6 (2010) 6.
[29] J. Ponomarenko, H.-H. Bui, W. Li, N. Fusseder, P.E. Bourne, A. Sette, B. Peters, ElliPro: a new structure-based tool for the prediction of antibody epitopes, BMC Bioinformatics 9 (2008) 514.

[30] S. Saha, G.P.S. Raghava, BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties, in: Artif. Immune Syst. Third Int. Conf. ICARIS 2004, Catania Sicily, Italy, Sept. 13-16, 2004 Proc, 2004, pp. 197–204.

[31] A.S. Kolaskar, P.C. Tongaonkar, A semi-empirical method for prediction of antigenic determinants on protein antigens, FEBS Lett. 276 (1990) 172–174.

[32] J.V. Kringelum, C. Lundegaard, O. Lund, M. Nielsen, Reliable B cell epitope predictions: impacts of method development and improved benchmarking, PLoS Comput. Biol. 8 (2012).

[33] S. Paul, E. Karosiene, S.K. Dhanda, V. Jurtz, L. Edwards, M. Nielsen, A. Sette, B. Peters, Determination of a predictive cleavage motif for eluted major histocompatibility complex class II ligands, Front. Immunol. 9 (2018) 1795.

[34] J. Laimer, H. Hofer, M. Fritz, S. Wegenkittl, P. Lackner, MAESTRO–multi agent stability prediction upon point mutations, BMC Bioinformatics 16 (2015) 116.

[35] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindah, Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1–2 (2015) 19–25.

[36] J. Huang, A.D. Mackerell, CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data, J. Comput. Chem. 34 (2013) 2135–2145.

[37] T.E.I.I.I. Cheatham, J.L. Miller, T. Fox, T.A. Darden, P.A. Kollman, Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins, J. Am. Chem. Soc. 117 (1995) 4193–4194.

[38] H.J.C. Berendsen, J.P.M. Postma, W.F. Van Gunsteren, A. Dinola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.

[39] Michael Holmboe, Import and plot Gromacs .xvg data files, in: Import Plot Gromacs .Xvg Data Files, 2020.

[40] N. Rapin, O. Lund, M. Bernaschi, F. Castiglione, Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system, PLoS One 5 (2010).

[41] A.A. Cruz, A. Souza-Machado, R. Franco, C. Souza-Machado, E.V. Ponte, P. M. Santos, M.L. Barreto, The impact of a program for control of asthma in a low-income setting, World Allergy Organ. J. 3 (2010) 167–174.

[42] T. Thalhamer, H. Dobias, T. Stepanoska, M. Pröll, H. Stutz, O. Dissertori, P. Lackner, F. Ferreira, M. Wallner, J. Thalhamer, A. Hartl, Designing hypoallergenic derivatives for allergy treatment by means of in silico mutation and screening, J. Allergy Clin. Immunol. 125 (2010).

[43] D. Martínez, M. Munera, J.F. Cantillo, J. Wortmann, J. Zakzuk, W. Keller, L. Caraballo, L. Puerta, An engineered hybrid protein from dermatophagoides pteronyssinus allergens shows hypoallergenicity, Int. J. Mol. Sci. 20 (2019).

[44] M. Curin, T. Garmatiuk, Y. Resch-Marat, K.W. Chen, G. Hofer, K. Fauland, W. Keller, W. Hemmer, S. Vrtala, M. Focke-Tejkl, R. Valenta, Similar localization of conformational IgE epitopes on the house dust mite allergens Der p 5 and Der p 21 despite limited IgE cross-reactivity, Allergy EurJ. Allergy Clin. Immunol. 73 (2018) 1653–1661.

[45] L. Aglas, A. Bethanis, P. Chrusciel, F. Stolz, M. Gruen, U.M. Jaakkola, L. Jongejan, E. Yatkin, R. Van Ree, In vivo induction of functional inhibitory IgG antibodies by a hypoallergenic Bet v 1 variant, Front. Immunol. 11 (2020) 1–15.

[46] X. Li, H.W. Yang, H. Chen, J. Wu, Y. Liu, J.F. Wei, In silico prediction of T and B cell epitopes of der f 25 in dermatophagoides farinae, Int. J. Genomics 2014 (2014).

[47] C.A. Akdis, Therapies for allergic inflammation: refining strategies to induce tolerance, Nat. Med. 18 (2012) 736–749.

[48] H. Chen, H.W. Yang, J.F. Wei, A.L. Tao, In silico prediction of the T-cell and IgE-binding epitopes of Per a 6 and Bla g 6 allergens in cockroaches, Mol. Med. Rep. 10 (2014) 2130–2136.

[49] G. Sircar, K. Jana, A. Dasgupta, S. Saha, S.G. Bhattacharya, Epitope mapping of rhi o 1 and generation of a hypoallergenic variant, J. Biol. Chem. 291 (2016) 18016–18029.

[50] H.C. da Santiago, S. Bennuru, J.M.C. Ribeiro, T.B. Nutman, Structural differences between human proteins and aero- and microbial allergens define allergenicity, PLoS One 7 (2012).

[51] W. Zhang, Y. Niu, Y. Xiong, M. Zhao, R. Yu, J. Liu, Computational prediction of conformational B-cell epitopes from antigen primary structures by ensemble learning, PLoS One 7 (2012) 1–9.

[52] S.L. Pang, K.L. Ho, J. Waterman, R.P. Rambo, A.H. Teh, I. Mathavan, G. Harris, K. Beis, Y.H. Say, M.S. Anusha, Y.Y. Sio, F.T. Chew, C.L. Ng, Crystal structure and epitope analysis of house dust mite allergen Der f 21, Sci. Rep. 9 (2019) 1–13.

[53] M. Wallner, M. Hauser, M. Himly, N. Zaborsky, S. Mutschlechner, A. Harrer, C. Asam, U. Pichler, R. Van Ree, P. Briza, J. Thalhamer, B. Bohle, G. Achatz, F. Ferreira, Reshaping the bet v 1 fold modulates TH polarization, J. Allergy Clin. Immunol. 127 (2011) 1–19.

[54] H. Hofer, C. Asam, M. Hauser, B. Nagl, J. Laimer, M. Himly, P. Briza, C. Ebner, R. Lang, T. Hawranek, B. Bohle, P. Lackner, F. Ferreira, M. Wallner, Tackling Bet v 1 and associated food allergies with a single hybrid protein, J. Allergy Clin. Immunol. 140 (2017) 525–533, e10.

[55] K.W. Chen, K. Blatt, W.R. Thomas, I. Swoboda, P. Valent, R. Valenta, S. Vrtala, Hypoallergenic Der p 1/Der p 2 combination vaccines for immunotherapy of house dust mite allergy, J. Allergy Clin. Immunol. 130 (2012) 435–443.e4.

[56] D. Zhang, R. Lazim, Application of conventional molecular dynamics simulation in evaluating the stability of apomyoglobin in urea solution, Sci. Rep. 7 (2017) 1–12.

[57] K. Condic-Jurkic, N. Subramanian, A.E. Mark, M.L. O'Mara, The reliability of molecular dynamics simulations of the multidrug transporter P-glycoprotein in a membrane environment, PLoS One 13 (2018) 1–24.

[58] M. Gill, M.E. McCully, Molecular dynamics simulations suggest stabilizing mutations in a de novo designed α/β protein, Protein Eng. Des. Sel. 32 (2019) 317–329.

[59] M.G. Pikkemaat, A.B.M. Linssen, H.J.C. Berendsen, D.B. Janssen, Molecular dynamics simulations as a tool for improving protein stability, Protein Eng. 15 (2002) 185–192.

[60] M. Garrido-Arandia, C. Gõmez-Casado, A. Díaz-Perales, L.F. Pacios, Molecular dynamics of major allergens from alternaria, birch pollen and peach, Mol. Inform. 33 (2014) 682–694.

[61] M. Hebditch, M.A. Carballo-Amador, S. Charonis, R. Curtis, J. Warwicker, Protein-Sol: a web tool for predicting protein solubility from sequence, Bioinformatics 33 (2017) 3098–3100.

[62] B.K. Bhandari, P.P. Gardner, in: TISIGNER Com : Web Services for Improving Recombinant Protein Production 49, 2021, pp. 654–661.

[63] A. Correa, P. Oppezzo, Tuning different expression parameters to achieve soluble recombinant proteins in E coli: advantages of high-throughput screening, Biotechnol. J. 6 (2011) 715–730.

[64] E. García-Fruitós, Insoluble proteins: methods and protocols, Insoluble Proteins Methods Protoc. 1258 (2014) 1–422.

[65] C. Gustafsson, J. Minshull, S. Govindarajan, J. Ness, A. Villalobos, M. Welch, Engineering genes for predictable protein expression, Protein Expr. Purif. 83 (2012) 37–46.

[66] G.L. Rosano, E.S. Morales, E.A. Ceccarelli, New tools for recombinant protein production in Escherichia coli: a 5-year update, Protein Sci. 28 (2019) 1412–1422.

[67] N. Zaborsky, M. Brunner, M. Wallner, M. Himly, T. Karl, R. Schwarzenbacher, F. Ferreira, G. Achatz, Antigen aggregation decides the fate of the allergic immune response, J. Immunol. 184 (2010) 725–735.

[68] N. Najafi, G. Hofer, P. Gattinger, D. Smiljkovic, K. Blatt, R. Selb, A. Stoecklinger, W. Keller, P. Valent, V. Niederberger, J. Thalhamer, R. Valenta, S. Flicker, Fusion proteins consisting of bet v 1 and phl p 5 form IgE-reactive aggregates with reduced allergenic activity, Sci. Rep. 9 (2019) 1–12.

[69] I.J. Ansotegui, G. Melioli, G.W. Canonica, R.M. Gómez, E. Jensen-Jarolim, M. Ebisawa, O. Luengo, L. Caraballo, G. Passalacqua, L.K. Poulsen, E. Savi, T. Zuberbier, E. Villa, J. Oppenheimer, R. Asero, J. Bernstein, J. Bousquet, V. Cardona, L. Cox, P. Demoly, et al., A WAO — ARIA — GA2LEN consensus document on molecular-based allergy diagnosis (PAMD@): update 2020, World Allergy Organ. J. 13 (2020), 100091.

[70] C.S. Pinheiro, E.S. Silva, E.M.M. de Andrade Belitardo, L.G.C. Pacheco, E.R.G. R. Aguiar, N.M. Alcantara-Neves, G. Gadermaier, F. Ferreira, En route to personalized medicine: uncovering distinct IgE reactivity pattern to house dust mite components in brazilian and austrian allergic patients, Clin. Transl. Allergy 11 (2021) 2–5.

[71] C.R. Kim, K.Y. Jeong, M.H. Yi, H.P. Kim, H.J. Shin, T.S. Yong, Cross-reactivity between group-5 and-21 mite allergens from Dermatophagoides farinae, Tyrophagus putrescentiae and Blomia tropicalis, Mol. Med. Rep. 12 (2015) 5467–5474.

[72] R. Valenta, B. Linhart, I. Swoboda, V. Niederberger, Recombinant allergens for allergen-specific immunotherapy: 10 years anniversary of immunotherapy with recombinant allergens, Allergy EurJ. Allergy Clin. Immunol. 66 (2011) 775–783.

[73] T.B. Casale, J.R. Stokes, Immunotherapy: what lies beyond, J. Allergy Clin. Immunol. 133 (2014) 612–619.

[74] A. Tscheppe, H. Breiteneder, Recombinant allergens in structural biology, diagnosis, and immunotherapy, Int. Arch. Allergy Immunol. 172 (2017) 187–202.

[75] R. Valenta, R. Campana, M. Focke-Tejkl, V. Niederberger, Vaccine development for allergen-specific immunotherapy based on recombinant allergens and synthetic allergen peptides: lessons from the past and novel mechanisms of action for the future, J. Allergy Clin. Immunol. 137 (2016) 351–357.

[76] S. Eichhorn, A. Hörschläger, M. Steiner, J. Laimer, B.M. Jensen, S.A. Versteeg, I. Pablos, P. Briza, L. Jongejan, N. Rigby, J.A. Asturias, A. Portolés, M. Fernandez-Rivas, N.G. Papadopoulos, A. Mari, L.K. Poulsen, P. Lackner, R. van Ree, F. Ferreira, G. Gadermaier, Rational design, structure-activity relationship, and immunogenicity of hypoallergenic pru p 3 variants, Mol. Nutr. Food Res. 63 (2019) 1–10.

[77] M.J. Robertson, J. Tirado-Rives, W.L. Jorgensen, Improved peptide and protein torsional energetics with the OPLS-AA force field, J. Chem. Theory Comput. 11 (2015) 3499–3509.

[78] S.L. Chan, T.C. Ong, Y.F. Gao, Y.S. Tiong, D.Y. Wang, F.T. Chew, Y.K. Mok, Nuclear magnetic resonance structure and IgE epitopes of Blo t 5, a major dust mite allergen, J. Immunol. 181 (2008) 2586–2596.

Received: 3 February 2021 | Revised: 2 July 2021 | Accepted: 6 August 2021

DOI: 10.1002/bit.27912

MINI-REVIEW

BIOTECHNOLOGY AND BIOENGINEERING WILEY

In vivo cleavage of solubility tags as a tool to enhance the levels of soluble recombinant proteins in *Escherichia coli*

Filipe S. R. Silva[1] | Sara P. O. Santos[1] | Roberto Meyer[1,2] | Eduardo S. Silva[1,2] | Carina S. Pinheiro[1,2] | Neuza M. Alcantara-Neves[1,2] | Luis G. C. Pacheco[1,3]

[1]Post-Graduate Program in Biotechnology, Institute of Health Sciences, Federal University of Bahia, Salvador, Bahia, Brazil

[2]Post-Graduate Program in Immunology, Institute of Health Sciences, Federal University of Bahia, Salvador, Bahia, Brazil

[3]Department of Biotechnology, Federal University of Bahia, Salvador, Bahia, Brazil

Correspondence

L. Pacheco, Post-Graduate Program in Biotechnology, Institute of Health Sciences, Federal University of Bahia, Salvador, BA, Brazil.
Email: luis.pacheco@ufba.br

Funding information

Fundação de Amparo à Pesquisa do Estado da Bahia, Grant/Award Number: PRONEM-2014

Abstract

Recombinant proteins are generally fused with solubility enhancer tags to improve the folding and solubility of the target protein of interest. However, the fusion protein strategy usually requires expensive proteases to perform in vitro proteolysis and additional chromatographic steps to obtain tag-free recombinant proteins. Expression systems based on intracellular processing of solubility tags in *Escherichia coli*, through co-expression of a site-specific protease, simplify the recombinant protein purification process, and promote the screening of molecules that fail to remain soluble after tag removal. High yields of soluble target proteins have already been achieved using these protease co-expression systems. Herein, we review approaches for controlled intracellular processing systems tailored to produce soluble untagged proteins in *E. coli*. We discuss the different genetic systems available for intracellular processing of recombinant proteins regarding system design features, advantages, and limitations of the various strategies.

KEYWORDS

recombinant proteins, controlled intracellular processing, *Escherichia coli*, protein solubility, site-specific protease

1 | INTRODUCTION

Escherichia coli is the microorganism of choice for the production of recombinant proteins (Feng et al., 2014). It is estimated that approximately 88% of protein structures deposited in the Protein Data Bank are derived from proteins produced in this host organism (Nettleship et al., 2010). In addition, this bacterial chassis produces more than 30% of FDA-approved biopharmaceuticals. The advantages of the *E. coli* expression systems include minimal requirements of laboratory structure and sterile procedures, short doubling time, high dry-weight yields in recombinant proteins, and straightforward process scale-up (Sezonov et al., 2007). However, a major bottleneck of this bacterial system involves the poor recovery of recombinant proteins in their soluble forms, particularly when working with proteins of eukaryotic origin (Costa et al., 2014).

To overcome solubility problems, a widely used approach is the fusion-tag technology, in which the gene encoding the target protein is fused with the coding sequence of a highly soluble protein. This technology often improves the solubility and stability of a given recombinant protein of interest, then contributing to a streamlined purification process (Kosobokova et al., 2016). The most commonly used solubility enhancers include Maltose-Binding Protein (MBP), Glutathione-S-Transferase (GST), Thioredoxin A (TrxA), and *N* Utilization Substance Protein A (NusA). The reasons why these proteins improve the solubilities of their partners are not completely understood to date; however, the current knowledge is that polypeptide chains rich in positively charged amino acids help to overcome aggregation by increasing electrostatic repulsion among residues during translation (Kang et al., 2015). Additionally, the tag-fusion technology can prevent defective mRNA structures when fused to the N-terminal portion of the target sequence. The solubility enhancer

proteins are also more likely to be highly soluble than shorter peptides (Waugh, 2005).

Nevertheless, the removal of fusion-protein tags requires expensive site-specific proteases, such as Tobacco Etch Virus Protease (TEVp), Human Rhinovirus Protease 3C (HRV 3C), Enterokinase, and Factor Xa. Moreover, additional chromatography steps are needed to remove the remaining solubility tags and the site-specific proteases from the final protein eluate (Li, 2011). In addition, some target proteins may still aggregate following the removal of the solubility tag. Therefore, in vivo intracellular processing can be useful as a tool to verify molecules that will fail to remain soluble after cleavage is performed (Cesaratto et al., 2016; Kapust & Waugh, 2000; Lu & Aon, 2014).

1.1 | In vivo cleavage systems for removal of solubility tags

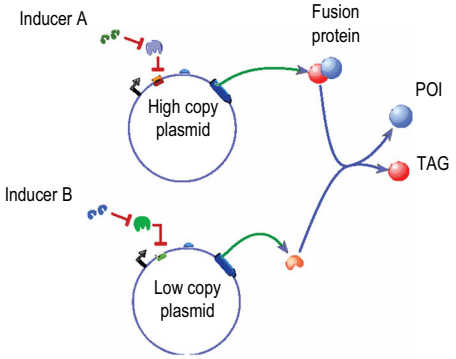
Even though in vivo experiments involving TEV protease had been reported back in the 90s (Parks et al., 1995), it was only in the 2000s that the co-expression of site-specific proteases was described as a strategy for permitting controlled intracellular processing (CIP) of recombinant proteins in *E. coli* (Kapust & Waugh, 2000) (Figure 1; Table 1). This foundational CIP system was based on the co-expression of a modified TEV protease along with a fusion target protein (MBP-GFP) containing a TEV recognition sequence for cleavage. This was a two-plasmid-based system, in which the gene cassette coding for the fusion protein of interest was physically segregated from the TEVp-encoding cassette, which was cloned in a separate low-copy number plasmid (Figure 1a) (Kapust & Waugh, 2000). Two different chemical inducers (isopropyl β-D-1-thiogalactopyranoside [IPTG] and tetracycline) were used to decouple the expression of the fusion protein from expression of the site-specific protease (Figure 1a). Expression of the TEVp was only induced by tetracycline after two hours of the fusion protein MBP-GFP; this rendered a significant improvement in the solubility of the passenger protein (Kapust & Waugh, 2000). The delayed induction of TEVp expression contributes to reduce metabolic burden in the host cell, which is one of the main factors affecting the yield and solubility of the target protein. Additionally, the use of a low copy number plasmid to drive protease expression avoids deleterious early metabolic burden in the cell (Figure 1a,b). By using this strategy, recombinant proteins that were formerly produced in an insoluble form in *E. coli* became soluble when fused with MBP and processed in vivo by TEVp. The first studies with controlled intracellular processing systems also demonstrated that this approach could be used as a diagnostic tool to determine if the passenger recombinant proteins would remain soluble when separated from their solubility tags (Kapust & Waugh, 2000).

Although double induction may contribute to reduce the metabolic burden, it may add an additional level of complexity to large-scale processes and then make it more expensive and time-consuming. A different approach was used to build another expression system capable of cleaving solubility tags in vivo and produce Diaminopropionate Ammonia-Lyase (DAL) in its soluble and functional form through co-expression of mutant variants of TEVp (Wei et al., 2012). In this system, the same promoter sequence was used for simultaneously expressing both the fusion protein and the protease by a single induction with IPTG (Figure 1c). This procedure could balance the side effects of using multiple inductions at different times, leading to a more straightforward process (Wei et al., 2012). The bacteria co-expressing the TEVpM2 variant effectively cleaved a GST-DAL fusion protein, and a higher yield of soluble and functional DAL was obtained in *E. coli* (Wei et al., 2012). Conversely, a significant decrease in cell growth was reported after only 5–6 h post-induction due to metabolic burden arising from overexpression of the two recombinant proteins (TEVp and DAL) at the same time. Additionally, it is also possible that using multiple *lac* operator sites will require more *LacI* repressor protein, which may lead to leakiness and anticipate metabolic stress.

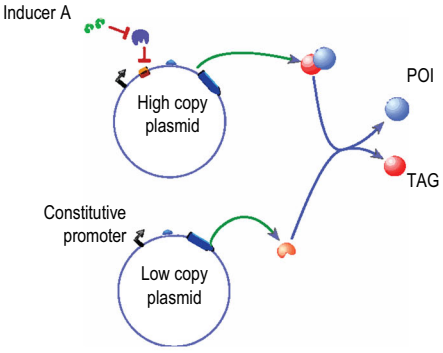
The use of genetic systems induced by physical rather than chemical stimuli may be another way to tackle the metabolic burden and to simplify the process of recombinant protein production. A temperature-sensitive pHsh promoter was used by Feng et al. (2014) to control the expression of the human rhinovirus protease 3C (HRV3C) in a CIP system in which the expression of the fusion protein of interest is controlled by a pT7-*lacO* promoter, under IPTG induction (Figure 1d). The pHsh promoter is activated by the alternative sigma factor 32 (σ^{32}) when the temperature is changed to 42°C. This physically induced CIP system successfully removed different protein tags (Trx, DsbA, GST, Nus, TF, and MBP) that were fused to the reporter protein Enhanced GFP (EGFP) (Table 1). The HRV3C was also fused to the GST tag to guarantee the solubility of the recombinant protease in vivo. The GST-HRV3C expression was induced by increasing temperature to 42°C for 1 h, whereas the target protein TAG-EGFP was induced for 5 h by IPTG. High levels of purified native EGFP were obtained with the Trx-tag fusion, following in vivo processing by HRV3C. The system was also successful with a difficult-to-express target, the Bluetongue virus (BTV) protein, which was produced in a soluble form when fused to a solubility-enhancer tag and digested by the HRV3C protease (Feng et al., 2014). Even though 42°C is not the ideal temperature for *E. coli* growth, the temperature-sensitive promoter could make the double induction protocol cheaper than the most commonly used chemical induction with different inducing agents. Therefore, these physically induced systems represent a good alternative to aid the identification of the best solubility-enhancer proteins and to scale-up the CIP process at a lower cost when compared with more traditional systems.

Although the foundational CIP systems effectively enhanced the solubilities of the target recombinant proteins, they generally required a two-plasmid-based system to drive expression of the fusion protein and the site-specific protease. Consequently, steps such as cloning, bacterial transformation, and antibiotic selection are generally performed twice, increasing method complexity. In an attempt to simplify cloning steps, a TEVp expression cassette was inserted into *the E. coli* chromosome by λ-Red recombineering

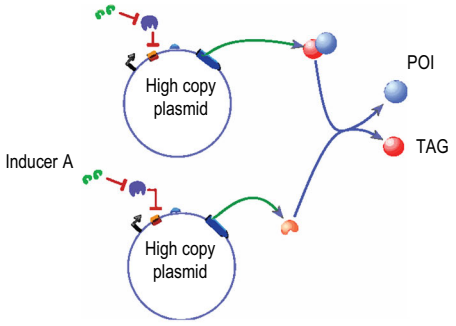
(a) **Temporal induction / Different inducers / Two plasmids** (Kapust & Waugh, 2000)



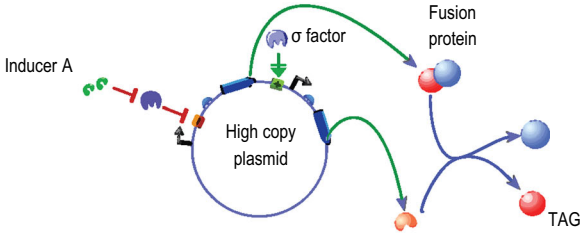
(b) **Inducible fusion protein / Constitutive protease / Two plasmids** (Donnelly et al., 2006)



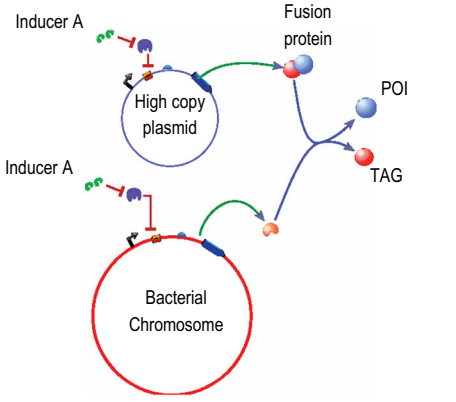
(c) **Simultaneous induction / Same inducer / Two plasmids** (Wei et al., 2012)



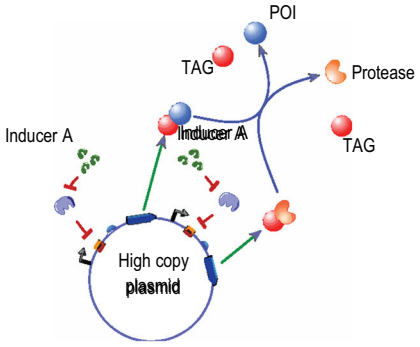
(d) **Temporal induction / Different stimuli / Single plasmid** (Feng et al., 2014)



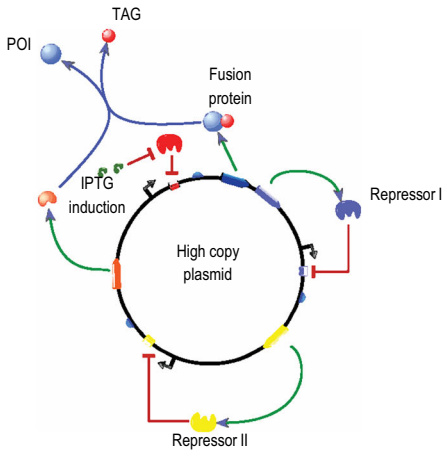
(e) **Simultaneous induction / Same inducer / Chromosomal protease expression** (Luo, Li, Chai, Zhang, & Shang, 2015)



(f) **Simultaneous induction / Same inducer / Single plasmid** (Kuo et al., 2011)



(g) **Self-regulated simultaneous induction / Single plasmid** (Silva et al., 2019)



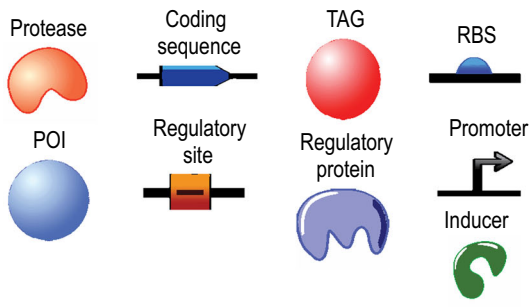


FIGURE 1 (See caption on next page)

(Luo et al., 2015) (Figure 1e). This was achieved by inserting a T7-TEV-*aacC1* into the chromosomal *malE* gene of *E. coli* to the cassette. The knock-in strain, termed *E. coli* LS2416, was then used to express GFP and N-TIMP (N-terminal inhibitory domain of human tissue inhibitor of metalloproteases-2), both fused to MBP. While the untagged version of N-TIMP was produced almost entirely in the insoluble form, the controlled intracellular processing within *E. coli* LS2416 rendered a high percentage of soluble recombinant proteins. The use of a single plasmid in this chromosome-based system may also contribute to reduce metabolic burden in the host cell, then enhancing expression of the target protein of interest (Shen et al., 2021).

1.2 | Optimizing proteases for in vivo proteolysis

Highly specific proteases are essential for building CIP systems. The purpose of using these enzymes is to eliminate the occurrence of nonspecific cleavage. Despite the low stability and solubility of wild-type TEV protease, this is the most co-expressed protease thanks to the high specificity of its target cleavage site ENLYFQG/S (Kapust et al., 2001; Parks et al., 1995). Wild-type TEVp may be problematic for in vivo processing though, due to low yield and to autoproteolysis that generates a truncated form with reduced activity (Kapust et al., 2001; Wei et al., 2012). Kapust et al. (2001) solved the autoproteolysis issue by mutating the amino acid sequences in the internal cleavage site of the enzyme, giving rise to a more stable and efficient TEVp (mutant S219V). Another alternative TEVp variant (TEV_{sh}) containing the mutations T17S/N68D/I77V was also described as being more soluble in vitro than natural TEVp (van den Berg et al., 2006). The mutant L56V/S135G, in turn, remained soluble at higher concentrations and displayed improved catalytic activity when compared with TEVp S219V (Cabrita et al., 2007). To achieve optimized in vivo proteolysis, Wei et al. (2012) combined all these aforementioned mutations in a new TEVp variant, termed TEVpM2. For this, the coding sequence for Emerald GFP (EmGFP) was cloned downstream of each TEVp variant, and then fluorescence analysis revealed that *E. coli* expressing the variant TEVpM2 (T17S/L56V/N68D/I77V/S135G) produced the highest fluorescence signal. This suggested that mutations in the TEVpM2 variant resulted in higher solubility in vivo.

Using a different approach, Nallamsetty et al. (2004) performed the cleavage of fusion proteins with the Tobacco Vein Mottling Virus protease (TVMVp), both in vivo and in vitro. TVMVp is active in a wide range of ionic strength, is highly active at low temperatures, and has catalytic efficiency that is comparable to TEVp. These two proteases display different sequence specificities and have high proteolytic stringency, so they rarely cut in nonspecific sites within the target protein. While the TEVp cleavage site is ENLYFQS, the canonical target site for TVMV is ETVRFQS. Therefore, TVMVp can replace TEVp when the fusion substrate has a peptide sequence that resembles the TEVp recognition site. Likewise, both proteases can be used together, allowing for the removal of two distinct tags (e.g., solubility tag and His-tag). Additionally, TVMVp has the advantage of not cleaving itself into inactive fragments, as wild-type TEVp does (Kapust et al., 2001).

The use of the Ubl-specific protease 1 (Ulp1) to cleave small ubiquitin modifying protein (SUMO) fusions has also been evaluated as an interesting alternative to site-specific proteases in controlled intracellular processing systems. Ulp1 is a highly specific protease that targets the tertiary structure of the SUMO tag for cleavage, rather than a specific stretch of amino acid residues within the protein; this permits cleavage without leaving any extra unwanted amino acid sequences in the terminus of the target protein (Butt et al., 2005). Moreover, the SUMO fusion tag exhibited increased solubility and stability during the expression of fused targets in *E. coli* (Kuo et al., 2011). Kuo and collaborators constructed a dual expression vector for co-expressing the His-tagged target protein GroQ and the Ulp1 protease, both fused to SUMO. Each fused sequence was inserted into pRSFDuetTM-1 (Novagen), which carries two multiple cloning sites, each one containing sequences to permit T7-regulated expression (Figure 1f) (Kuo et al., 2011). The simultaneous expressions of SUMO-Ulp1 and SUMO-fused protein resulted in the production of the protein of interest GroQ in its native form, as confirmed by Ni-NTA chromatography and sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The yields of SUMO-GroQ (without co-expressing Ulp1) and processed GroQ were comparable after standard purification, indicating that the GroQ

FIGURE 1 Schematic of controlled intracellular processing (CIP) systems in *Escherichia coli*. (a) Double induction at different times using an inducer molecule (e.g., IPTG) for activating fusion protein production and a second inducer (e.g., Anhydrotetracycline or Arabinose) for stimulating protease expression. (b) Expression of the fusion protein of interest (POI) is regulated by a chemically inducible promoter, whereas the site-specific protease is constitutively expressed from a low copy number plasmid. (c) Simultaneous double induction of fusion protein and site-specific protease using the same inducer molecule (*e.g.*, IPTG). (d) Target fusion protein is chemically induced (*e.g.*, IPTG), and protease expression is activated under stress condition by sigma (σ) transcription factor (e.g., turning the temperature to 42°C). (e) The bacterial strain contains a protease expression unit in its chromosome to produce the protease, and the fusion protein is expressed from a plasmid. (f) Both target fusion protein and protease are subcloned in distinct multiple cloning sites of a single plasmid, under the control of the same T7 promoter. (g) The production of both POI and protease are activated by the induction of a single promoter by IPTG. IPTG addition to the media generates the translation of the fusion protein and the first repressor protein. The first repressor inhibits production of the second repressor protein through binding to the promoter's operator site. This releases expression of the site-specific protease to perform cleavage of the fusion protein. IPTG, isopropyl β-D-1-thiogalactopyranoside; RBS, ribosome binding sites

TABLE 1 Overview of studies that have used controlled intracellular processing for recombinant protein production

	Target protein	Protease induction	Additional information	Refs.
TEV protease				
MBP	Green fluorescent protein (GFP), TIMP *N*-terminal inhibitory domain of human tissue inhibitor of metalloproteinases-2 (TIMP), Human cyclin-dependent kinase 4 inhibitor (p16), and oncoprotein encoded by human papillomavirus (E6)	Anhydrotetracycline	The fusion protein was almost entirely processed when a 2-h delayed induction of the protease was applied	Kapust and Waugh (2000)
	Human IL-13 (hIL-13)	Constitutive	The yield of the in vivo processed proteins was superior when compared with in vitro processing. ca. 11 mg of proteins per liter of culture were obtained	Eisenmesser et al. (2000)
	S-adenosylho-mocysteine hydrolase	Constitutive	The proteins were purified using affinity chromatography; purified fractions only contained the processed target protein	Bujnicki et al., (2003)
	Enoyl reductase enzyme (ENR)	Constitutive	After subtractive amylose column purification, pure target fractions were concentrated to 20 mg ml^{-1}	Muench et al. (2006)
	Type III secretion system effector (YopR)	Anhydrotetracycline	SDS-PAGE and mass spectrometry confirmed that YopR was obtained in its full-length	Schubot et al., (2005)
	Enoyl reductase (ENR)	Constitutive	An EtENR concentration of 33 mg/ml was obtained after diafiltration with a 30 kDa concentrator	J. Z. Lu et al. (2007)
	Yersinia modulating protein (YmoA)	Constitutive	The fusion protein was cleaved in vivo to generate YmoA with a C-terminal His-tag	McFeeters et al. (2007)
	Escherichia coli lipoate ligase (EcLplA); lipoate ligase 1 (LipL1); - lipoate ligase 2 (LipL2)	Constitutive	After metal chelate chromatography, an additional cation exchange chromatography was needed for obtaining untagged proteins	Afanador et al. (2014)
	dCAS9	Isopropyl β-D-1-thiogalactopyranoside	Co-expression of MBP-ENLYFQG-dCas9 fusion, TEVp, and specific sgRNA targeting a promoter, resulted in the tight regulation of GFP in *E. coli*	Didovyk et al. (2016)
	Green fluorescent protein (GFP), *N*-terminal inhibitory domain of human tissue inhibitor of metalloproteases-2 (TIMP)	Isopropyl β-D-1-thiogalactopyranoside	The proportion of released purified proteins estimated by SDS-PAGE was higher than 90%	Luo et al. (2015)
MBP and NusA	Green fluorescent protein (GFP), glyceraldehyde 3-phosphate dehydrogenase (G3PDH), dihydrofolate reductase (DHFR), rhodanese, luciferase, tissue	Constitutive	MBP fused targets were more soluble after in vivo processing by TEVp than NusA fused targets	Nallamsetty and Waugh (2006)

(Continues)

TABLE 1 (Continued)

	Target protein	Protease induction	Additional information	Refs.
	inhibitor of metalloproteinases-1 (TIMP), YopN, YopJ, YopT, YscK, YscL, and YscO			
Acyl carrier protein (ACP)	Glucokinase (GlcK), α-amylase (Amy), and GFP	Arabinose (0.2%) or IPTG (0.4 mM)	Under co-expression with TEVp, target proteins were separated conveniently by one-step IMAC, and retained their enzymatic activities	Wang et al. (2015)
GST	Diaminopropionate ammonia-lyase (DAL), maize 2-Cys peroxiredoxin A (Prx)	Isopropyl β-D-1-thiogalactopyranoside	Target proteins cleaved in vivo by TEVpM2 variants displayed increased enzymatic activity	Wei et al. (2012)
KDPG Aldolase (EDA)	Enhanced green fluorescent protein (EGFP)	Isopropyl β-D-1-thiogalactopyranoside	Total soluble recombinant protein yield of 272.0 ± 60.1 μg per ml of culture after diafiltration	Silva et al. (2019)
TVMV protease				
MBP, GST, and TRX.	Transcription termination/antitermination protein NusG	Anhydrotetracycline	Intracellular processing of virtually all fusion protein substrates was achieved by using TVMV protease	Nallamsetty et al., 2004)
MBP	Hypothetical (*B. cereus*), Cytoplasmic protein, Regulatory protein, RNA ligase, Cytoplasmic protein Hypothetical, Inner membrane proteinC, Inner membrane protein, Cytoplasmic protein C, Hydrophilic protein, Regulatory proteinC, SAM methyltransferaseC, Galactitol enzyme IIA, Transport protein, Hypothetical, Urease accessory protein	Constitutive	The average yield using this system was 39 mg per liter of culture	Donnelly et al. (2006)
Human rhinovirus 3C (HRV3C) protease				
MBP	Super folder green fluorescent protein sf-GFP	L-arabinose (0.2%)	A maximum concentration of 5 mg per gram of cell weight for the purified target proteins was produced	Raran-Kurussi and Waugh (2016)

TABLE 1 (Continued)

	Target protein	Protease induction	Additional information	Refs.
TRX, DsbA, GST, Nus, TF, and MBP	EGFP, Bluetongue virus (BTV) protein	Temperature (42°C)	The amount of free EGFP in SDS-PAGE measured with ImageJ was higher when expressed fused with the TRX solubility tag	Feng et al., (2014)
Ubl-specific protease 1				
SUMO	Q domain of Drosophila Groucho (GroQ)	Isopropyl β-D-1-thiogalactopyranoside	Simultaneous induction of SUMO-GroQ and SUMO-Ulp1 produced approximately 1 mg of purified GroQ per gram of wet *E. coli*	Kuo et al. (2011)

Abbreviations: MBP, maltose-binding protein; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; TEV, Tobacco Etch Virus.

protein retained its soluble form after the in vivo cleavage and purification (Kuo et al., 2011).

1.3 | CIP systems can facilitate protein purification

There are several well-established, straightforward, semi-automated protocols for high throughput protein purification. However, most of them are adapted for His-tagged proteins or fully native proteins. To purify recombinant proteins that are fused to solubility tags, two consecutive steps of affinity chromatography are often necessary. The first chromatographic procedure purifies the fusion protein and, following site-specific proteolysis, a second subtractive chromatographic procedure is generally implemented to remove the cleaved tag and the site-specific protease (Wang et al., 2015). Donnelly and collaborators used the CIP approach to obtain untagged proteins without depending on a second chromatographic step (Donnelly et al., 2006). They took advantage of the TEVp and TVMVp distinct target specificities and modified fusion recombinant protein to contain two distinct protease recognition sites (MBP–TVMVsite–His6-tag–TEVsite-target protein). This strategy was innovative in comparison to most CIP systems, as the His6-tag can also be removed from the target protein. The authors co-expressed TVMVp with 16 difficult-to-purify target proteins, that were fused to MBP. TVMVp was produced constitutively in the BL21 strain, and the release of the MBP tag was confirmed by SDS-PAGE in all experiments, demonstrating that all recombinant fusion proteins were cleaved by TVMVp through in vivo processing (Figure 1b). Based on the abundance of proteins recovered in the soluble fraction, 10 out of 16 proteins were considered as sufficiently soluble, and two of them were found in both soluble and insoluble fractions. These findings confirm that CIP systems can be used for early screening of proteins that will become insoluble when separated from their solubility partners in vitro. The remaining proteins were purified by efficient standard protocols for his-tagged proteins. This dual tag approach proves that using in vivo proteolysis can improve purity and yield in semi-automated protocols.

1.4 | Improving the solubility of target proteins using synthetic genetic circuits

In a recent study, our group demonstrated that a genetic regulatory cascade could be used to control the in vivo removal of a solubility tag from a fusion recombinant protein using a single plasmid (Silva et al., 2019) (Figure 1g). To assemble this expression vector, termed pSOLC, the genetic modules were built as follow: (i) a first module contained the coding sequence for the solubility tag KDPG aldolase (EDA), a Gly-Ser-Gly-Ser flexible linker, and the canonical TEVp cleavage recognition site; (ii) the second module encoded the target protein EGFP and was placed under the control of the module #1; and (iii) the third module was designed to express the TetR repressor and TEVp.

Significant metabolic burden may arise from the expression of the TEV protease and the regulatory proteins, in addition to the fusion target protein. To reduce this metabolic load, the system was designed using promoters and ribosome binding sites (RBS) of different strengths; additionally, we introduced LVA degradation tails in the regulatory proteins (TetR and cI434) to expedite degradation and prevent the circuit from collapsing due to the accumulation of regulators. Although the levels of fusion target protein obtained in a two-plasmid system configuration were not significantly different from the yields obtained in the single plasmid configuration, the two-plasmid-based system mostly rendered the target protein still in its fused form, indicating poor production of the protease. This might be due to augmented metabolic load in the two-plasmid-based system caused by replication of a second plasmid with an additional antibiotic resistance marker.

The single plasmid system produced a total soluble recombinant protein yield of 272.0 ± 60.1 μg/ml of culture. In addition, free EGFP composed an average 46.5% (max. 67%) of the total purified protein fraction and was easily separated from the remaining EDA-EGFP. The advantage of this regulatory cascade is the interaction of genetic elements by a single chemical input, leading to the simultaneous production of a fusion recombinant protein and a site-specific protease, which then cleaves the solubility tag from the target protein. However, a significant part of the recombinant protein remained in its fusion form after intracellular processing, which could be a limitation of the approach. The strategy merits further developments due to the requirement of a single induction with only one inducer as well as the possibility of use in different cell lineages (Silva et al., 2019).

2 | CONCLUSION

It is unlikely that CIP systems will replace standard expression systems for the production of fused recombinant proteins, but it is worth considering them as an alternative to solve problems related to protein solubility and purification methodology. Three main challenges may hamper the applicability of CIP systems in recombinant protein production: (i) co-expression of proteases may lead to metabolic burden and, consequently, poor soluble protein yield; (ii) using multiple inductions could make target protein production more expensive and complicated; and (iii) the need for several cloning steps can be time and resource-consuming in some genetic configurations. As discussed herein, the overexpression of two heterologous proteins at the same level by the bacterial cells, using strong regulatory sequences, will intensify the metabolic load and may lead to poor yield or insoluble production of the recombinant protein of interest. Stress-responsive promoters might be an alternative to avoid the metabolic burden and make the process cheaper and less complicated. Redesigning the CIP genetic systems in a genetic circuit format may help to balance the proteins' stoichiometries and prevent the fast drain of the host´s resources.

ACKNOWLEDGEMENTS

FSRS is the recipient of a scholarship from FAPESB. LGCP and NMAN are recipients of research fellowships from the National Council for Scientific and Technological Development of Brazil (CNPq). Work at our group is partially funded through the FAPESB/CNPq PRONEM-2014 grant awarded to LGCP.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Filipe S. R. Silva and Luis G. Carina S. Pinheiro conceived the manuscript idea. Filipe S. R. Silva led the writing and construction of figure. Sara P. O. Santos and Eduardo S. Silvaara P. O. Santos contributed with literature search and manuscript writing. Roberto Meyer, Carina S. Pinheiro, Neuza M. Alcantara-Neves, and Luis G. Carina S. Pinheiro reviewed the draft versions, read, and approved the final manuscript.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES

Afanador, G. A., Matthews, K. A., Bartee, D., Gisselberg, J. E., Walters, M. S., Freel Meyers, C. L., & Prigge, S. T. (2014). Redox-dependent lipoylation of mitochondrial proteins in Plasmodium falciparum. *Molecular Microbiology*, *94*(1), 156–171. https://doi.org/10.1111/mmi.12753

Bujnicki, J. M., Prigge, S. T., Caridha, D., & Chiang, P. K. (2003). Structure, evolution, and inhibitor interaction of S-adenosyl-L-homocysteine hydrolase from Plasmodium falciparum. *Proteins*, *52*(4), 624–632. https://doi.org/10.1002/prot.10446

van den Berg, S., Lofdahl, P. A., Hard, T., & Berglund, H. (2006). Improved solubility of TEV protease by directed evolution. *Journal of Biotechnology*, *121*(3), 291–298. https://doi.org/10.1016/j.jbiotec.2005.08.006

Butt, T. R., Edavettal, S. C., Hall, J. P., & Mattern, M. R. (2005). SUMO fusion technology for difficult-to-express proteins. *Protein Expression and Purification*, *43*(1), 1–9. https://doi.org/10.1016/j.pep.2005.03.016

Cabrita, L. D., Gilis, D., Robertson, A. L., Dehouck, Y., Rooman, M., & Bottomley, S. P. (2007). Enhancing the stability and solubility of TEV protease using in silico design. *Protein Science*, *16*(11), 2360–2367. https://doi.org/10.1110/ps.072822507

Cesaratto, F., Burrone, O. R., & Petris, G. (2016). Tobacco Etch virus protease: A shortcut across biotechnologies. *Journal of Biotechnology*, *231*, 239–249. https://doi.org/10.1016/j.jbiotec.2016.06.012

Costa, S., Almeida, A., Castro, A., & Domingues, L. (2014). Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: The novel Fh8 system. *Frontiers in Microbiology*, *5*, 63. https://doi.org/10.3389/fmicb.2014.00063

Didovyk, A., Borek, B., Hasty, J., & Tsimring, L. (2016). Orthogonal Modular Gene Repression in Escherichia coli Using Engineered CRISPR/Cas9. *ACS Synthetic Biology*, *5*(1), 81–88. https://doi.org/10.1021/acssynbio.5b00147

Donnelly, M. I., Zhou, M., Millard, C. S., Clancy, S., Stols, L., Eschenfeldt, W. H., Collart, F. R., & Joachimiak, A. (2006). An expression vector tailored for large-scale, high-throughput

purification of recombinant proteins. *Protein Expression and Purification*, *47*(2), 446–454. https://doi.org/10.1016/j.pep.2005.12.011

Eisenmesser, E. Z., Kapust, R. B., Nawrocki, J. P., Mazzulla, M. J., Pannell, L. K., Waugh, D. S., & Byrd, R. A. (2000). Expression, purification, refolding, and characterization of recombinant human interleukin-13: utilization of intracellular processing. *Protein Expression and Purification*, *20*(2), 186–195. https://doi.org/10.1006/prep.2000.1283

Feng, Y., Xu, Q., Yang, T., Sun, E., Li, J., Shi, D., & Wu, D. (2014). A novel self-cleavage system for production of soluble recombinant protein in *Escherichia coli*. *Protein Expression and Purification*, *99*, 64–69. https://doi.org/10.1016/j.pep.2014.04.001

Kang, Y. S., Song, J. A., Han, K. Y., & Lee, J. (2015). *Escherichia coli* EDA is a novel fusion expression partner to improve solubility of aggregation-prone heterologous proteins. *Journal of Biotechnology*, *194*, 39–47. https://doi.org/10.1016/j.jbiotec.2014.11.025

Kapust, R. B., Tozser, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D., & Waugh, D. S. (2001). Tobacco Etch virus protease: Mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Engineering*, *14*(12), 993–1000. https://doi.org/10.1093/protein/14.12.993

Kapust, R. B., & Waugh, D. S. (2000). Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expression and Purification*, *19*(2), 312–318. https://doi.org/10.1006/prep.2000.1251

Kosobokova, E. N., Skrypnik, K. A., & Kosorukov, V. S. (2016). Overview of fusion tags for recombinant proteins. *Biochemistry*, *81*(3), 187–200. https://doi.org/10.1134/S0006297916030019

Kuo, D., Nie, M., De Hoff, P., Chambers, M., Phillips, M., Hirsch, A. M., & Courey, A. J. (2011). A SUMO-Groucho Q domain fusion protein: Characterization and in vivo Ulp1-mediated cleavage. *Protein Expression and Purification*, *76*(1), 65–71. https://doi.org/10.1016/j.pep.2010.08.008

Li, Y. (2011). Self-cleaving fusion tags for recombinant protein production. *Biotechnology Letters*, *33*(5), 869–881. https://doi.org/10.1007/s10529-011-0533-8

Lu, Q., & Aon, J. C. (2014). Co-expression for intracellular processing in microbial protein production. *Biotechnology Letters*, *36*(3), 427–441. https://doi.org/10.1007/s10529-013-1379-z

Lu, J. Z., Muench, S. P., Allary, M., Campbell, S., Roberts, C. W., Mui, E., McLeod, R. L., Rice, D. W., & Prigge, S. T. (2007). Type I and type II fatty acid biosynthesis in Eimeria tenella: enoyl reductase activity and structure. *Parasitology*, *134*(Pt.14), 1949–1962. https://doi.org/10.1017/s0031182007003319

Luo, X., Li, L., Chai, M., Zhang, Q., & Shang, G. (2015). Escherichia coli BL21(DE3) chromosome-based controlled intracellular processing system for fusion protein separation. *Journal of Microbiological Methods*, *114*, 35–37. https://doi.org/10.1016/j.mimet.2015.04.013

McFeeters, R. L., Altieri, A. S., Cherry, S., Tropea, J. E., Waugh, D. S., & Byrd, R. A. (2007). The High-Precision Solution Structure of Yersinia Modulating Protein YmoA Provides Insight into Interaction with H-NS. *Biochemistry*, *46*(49), 13975–13982. https://doi.org/10.1021/bi701210j

Muench, S. P., Prigge, S. T., Zhu, L., Kirisits, M. J., Roberts, C. W., Wernimont, S., McLeod, R., & Rice, D. W. (2006). Expression, purification and preliminary crystallographic analysis of the Toxoplasma gondii enoyl reductase. *Acta Crystallographica Section F*, *62*(6), 604–606. https://doi.org/10.1107/S1744309106018112

Nallamsetty, S., Kapust, R. B., Tozser, J., Cherry, S., Tropea, J. E., Copeland, T. D., & Waugh, D. S. (2004). Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro. *Protein Expression and Purification*, *38*(1), 108–115. https://doi.org/10.1016/j.pep.2004.08.016

Nallamsetty, S., & Waugh, D. S. (2006). Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expression and Purification*, *45*(1), 175–182. https://doi.org/10.1016/j.pep.2005.06.012

Nettleship, J. E., Assenberg, R., Diprose, J. M., Rahman-Huq, N., & Owens, R. J. (2010). Recent advances in the production of proteins in insect and mammalian cells for structural biology. *Journal of Structural Biology*, *172*(1), 55–65. https://doi.org/10.1016/j.jsb.2010.02.006

Parks, T. D., Howard, E. D., Wolpert, T. J., Arp, D. J., & Dougherty, W. G. (1995). Expression and purification of a recombinant tobacco etch virus NIa proteinase: Biochemical analyses of the full-length and a naturally occurring truncated proteinase form. *Virology*, *210*(1), 194–201. https://doi.org/10.1006/viro.1995.1331

Raran-Kurussi, S., & Waugh, D. S. (2016). A dual protease approach for expression and affinity purification of recombinant proteins. *Analytical Biochemistry*, *504*, 30–37. https://doi.org/10.1016/j.ab.2016.04.006

Schubot, F. D., Cherry, S., Austin, B. P., Tropea, J. E., & Waugh, D. S. (2005). Crystal structure of the protease-resistant core domain of Yersinia pestis virulence factor YopR. *Protein Sci*, *14*(6), 1679–1683. https://doi.org/10.1110/ps.051446405

Sezonov, G., Joseleau-Petit, D., & D'Ari, R. (2007). *Escherichia coli* physiology in Luria-Bertani broth. *Journal of Bacteriology*, *189*(23), 8746–8749. https://doi.org/10.1128/JB.01368-07

Shen, Q., Yu, Z., Zhou, X. -T., Zhang, S. -J., Zou, S. -P., Xiong, N., Xue, Y. P., Liu, Z. Q., & Zheng, Y. G. (2021). Identification of a novel promoter for driving antibiotic-resistant genes to reduce the metabolic burden during protein expression and effectively select multiple integrations in Pichia Pastoris. *Applied Microbiology and Biotechnology*, *105*(8), 3211–3223. https://doi.org/10.1007/s00253-021-11195-0

Silva, F. S. R., Santos, S. P. O., Meyer, R., Alcantara-Neves, N. M., Pinheiro, C. S., & Pacheco, L. G. C. (2019). Single-input regulatory cascade for in vivo removal of the solubility tag in fusion recombinant proteins produced by *Escherichia coli*. *Frontiers in Bioengineering and Biotechnology*, *7*, 200. https://doi.org/10.3389/fbioe.2019.00200

van den Berg, S., Lofdahl, P. A., Hard, T., & Berglund, H. (2006). Improved solubility of TEV protease by directed evolution. *J Biotechnol*, *121*(3), 291–298. https://doi.org/10.1016/j.jbiotec.2005.08.006

Wang, H. Z., Chu, Z. Z., Chen, C. C., Cao, A. C., Tong, X., Ouyang, C. B., Yuan, Q. H., Wang, M. N., Wu, Z. K., Wang, H. H., & Wang, S. B. (2015). Recombinant passenger proteins can be conveniently purified by one-step affinity chromatography. *PLoS One*, *10*(12), e0143598. https://doi.org/10.1371/journal.pone.0143598

Waugh, D. S. (2005). Making the most of affinity tags. *Trends in Biotechnology*, *23*(6), 316–320. https://doi.org/10.1016/j.tibtech.2005.03.012

Wei, L., Cai, X., Qi, Z., Rong, L., Cheng, B., & Fan, J. (2012). In vivo and in vitro characterization of TEV protease mutants. *Protein Expression and Purification*, *83*(2), 157–163. https://doi.org/10.1016/j.pep.2012.03.011

How to cite this article: Silva, F. S. R., Santos, S. P. O., Meyer, R., Silva, E. S., Pinheiro, C. S., Alcantara-Neves, N. M., & Pacheco, L. G. C. (2021). In vivo cleavage of solubility tags as a tool to enhance the levels of soluble recombinant proteins in *Escherichia coli*. *Biotechnology and Bioengineering*, 118, 4159–4167. https://doi.org/10.1002/bit.27912

Vaccine 38 (2020) 4762–4772

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Immunogenicity and protection induced by recombinant *Toxocara canis* proteins in a murine model of toxocariasis

Luis Fabián Salazar Garcés [a], Leonardo Freire Santiago [a], Sara Patrícia de Oliveira Santos [a], Dumar Alexander Jaramillo Hernández [a,c], Marcia Barbosa da Silva [a], Vitor dos Santos Alves [a], Elisania Fontes Silveira [a], Stella Maria Barrouin-Melo [b], Philip John Cooper [d,e], Luis Gustavo Carvalho Pacheco [a], Carina da Silva Pinheiro [a], Neuza Maria Alcantara-Neves [a,*]

[a] Institute of Health Sciences, Federal University of Bahia, Salvador, Bahia, Brazil
[b] Department of Veterinary Anatomy, Pathology and Clinics, School of Veterinary Medicine and Zootechny, Federal University of Bahia, Salvador, Bahia 40170-110, Brazil
[c] Animal Sciences College, University of the Llanos, Villavicencio, Meta, Colombia
[d] Institute of Infection and Immunity, St George's University of London, London, United Kingdom
[e] Facultad de Ciencias Medicas, de la Salud y la Vida, Universidad Internacional del Ecuador, Quito, Ecuador

ARTICLE INFO

Article history:
Received 11 November 2019
Received in revised form 24 April 2020
Accepted 27 April 2020
Available online 23 May 2020

Keywords:
Toxocara canis
Recombinant proteins
In silico analysis
Vaccine candidates

ABSTRACT

Toxocariasis, a natural helminth infection of dogs and cats caused by *Toxocara canis* and *T. cati*, respectively, that are transmitted to mammals, including humans. Infection control is based currently on periodic antihelmintic treatment and there is a need for the development of vaccines to prevent this infection. Materials and Methods: Eight potential vaccine candidate *T. canis* recombinant proteins were identified by *in silico* (rTcGPRs, rTcCad, rTcVcan, rTcCyst) and larval proteomics (rTES26, rTES32, rMUC-3 and rCTL-4) analyses. Immunogenicity and protection against infectious challenge for seven of these antigens were determined in a murine model of toxocariasis. C57BL/6 female mice were immunized with each of or combinations of recombinant antigens prior to challenge with 500 *T. canis* embryonated eggs. Levels of specific antibodies (IgG, IgG1, IgG2a and IgE) in sera and cytokines (IL-5, INF-ɣ and IL-10) produced by antigens-stimulated splenocytes, were measured. Presence of specific antibodies to the molecules was measured in sera of *T. canis*-seropositive dogs and humans. Results: All seven molecules were immunogenic in immunized mice; all stimulated significantly elevated levels of specific IgG, IgG1 or IgG2a and six were associated with elevated levels of specific IgE; all induced elevated production of IFN- ɣ and IL-10 by splenocytes, but only the *in silico*-identified membrane-associated recombinants (rTcCad, rTcVcan, and rTcCyst) induced significantly increased IL-5 production. Vaccination with two of the latter (rTcCad and rTcVcan) reduced larval loads in the *T. canis* challenged mice by 54.3% and 53.9% ($P < 0.0001$), respectively, compared to unimmunized controls. All seven recombinants were recognized by *T. canis*-seropositive dog and human sera. Conclusion: The identification of vaccine targets by *in silico* analysis was an effective strategy to identify immunogenic *T. canis* proteins capable of reducing larval burdens following challenge with the parasite. Two recombinant proteins, rTcCad and rTcVcan, were identified as promising vaccine candidates for canine toxocariasis.

Abbreviations: A. caninum, *Ancylostoma caninum*; CEUA, Animal Experimentation and Use Commission; CNPq, National Council for Scientific and Technological Development; FAPESB, Research's Foundation of the State of Bahia; FIP, Fungi Lipase-International FIP Standard; ICS, Institute of Health Sciences; PSORT, Program for predicting protein localization in cells; rCTL4, Recombinant antigen of CTL4; rMUC3, Recombinant antigen of MUC3; rTcCad, Recombinant protein homologates to cadherin; rTcCyst, Recombinant protein homologates cystatin; rTcGPRs, Recombinant protein homologates to G protein coupled receptor; rTcVcan, Recombinant protein homologates to potassium channels; *S. lupi*, *Spirocerca lupi*; SOSUI, Program for classification and prediction of secondary structure of membrane proteins; SignalP, Bioinformatics program that identifies signal peptides; *T. canis*, *Toxocara canis*; *T. cati*, *Toxocara cati*; *T. vitulorum*, *Toxocara vitulorum*; TES-32, CTL-1 - Type 1 Type-C Lectins; TMHMM, Program for predicting transmembrane propellers in proteins; UFBA, University Federal of Bahia.

* Corresponding author at: Laboratory of Allergy and Acarology, Institute of Health Sciences, Federal University of Bahia, Avenida Rector Miguel Calmon, without n°, Vale do Canela, 40110-100 Salvador, Bahia, Brazil.

E-mail addresses: dumar.jaramillo@unillanos.edu.co (D.A. Jaramillo Hernández), barrouin@ufba.br (S.M. Barrouin-Melo), pcooper@sgul.ac.uk (P.J. Cooper), neuzalcantara@gmail.com (N.M. Alcantara-Neves).

https://doi.org/10.1016/j.vaccine.2020.04.072
0264-410X/

1. Introduction

Infections with *Toxocara canis* and *T. cati* in definitive animal hosts (i.e. dogs and cats) can occur during lactation, through ingestion of embryonated eggs in the environment, or through ingestion of larvae present in tissues of paratenic hosts [1,2]. Furthermore, intrauterine transmission of *T. canis* from infected pregnant dogs to their offspring, through migration of larvae via the placenta, is considered the most important route of transmission of infection in dogs [3]. The infection has a worldwide distribution with variable prevalence rates reported among dogs and cats in European countries ranging from 3.5% to 34% [4–8]. Higher prevalence is generally reported in tropical and subtropical settings [9,10]. The presence of *T. canis* eggs in dog stool samples in Latin America is reported to vary considerably, from 8.7% in São Paulo, Brazil [11], to 16.5% in the Galápagos Islands, Ecuador [12], to 43.6% in Colombia [9]. A study from Salvador, Brazil, estimated 85% of dogs to be seropositive for *Toxocara* spp. [10].

Toxocara spp. are able to modulate the host immune response to favor the long-term survival of parasites within the host through the release of larval immune modulatory excreted/secreted products (TES). TES are derived from the epicuticle surface coating that is rich in molecules such as mucin and type C lectins [13,14], as documented by genomic and proteomic studies [15,16].

Previous studies, using TES soluble products, irradiated eggs of *T. canis*, or soluble larval and adult somatic antigens of *T. vitulorum*, were able to induce immune protection against these parasites in murine models [17]. The most likely candidate vaccine antigens are considered to be those associated with the larval surface of the parasite [18]. Molecules capable of generating a protective immune response may be present also in TES antigens of infective larvae that include 'hidden' somatic antigens [19].

The development of effective vaccines against parasite infections although challenging, provide a promising alternative to traditional drug-based treatments. For toxocariasis, the immunization of dogs and cats could block larval re-activation and transmission from the mothers to their litters through intrauterine or breastfeeding routes. Effective control of transmission from these animals likely will contribute to reduced environmental contamination and prevent infection of natural and paratenic hosts including humans. Practical barriers to vaccine development include difficulty in obtaining sufficient quantities of parasite antigens and their preparation into stable and standardized formulations required for commercialization. For these reasons, the use of recombinant molecules provides an attractive solution [20].

In silico analyses allow the identification of novel vaccine antigens including "hidden antigens". The latter represent molecules to which the host immune system is not usually exposed. Specific antibodies generated against these antigens are capable of neutralizing biological or metabolic activities of the parasites, leading to their death and/or expulsion from the host [21–23]. This rationale has been used to develop highly effective recombinant antigen-based vaccines against *Echinococcus granulosus* and *Taenia ovis* in sheep, and *Taenia saginata* in cattle [24–26].

In the present study, we used *in silico* and proteomics analyses to identify protein antigens derived from the surface or TES of *T. canis* larvae as potential vaccine candidates against *T. canis* infection.

2. Materials and methods

2.1. Identification and in silico selection of potential vaccine targets

Fig. 1 shows *T. canis* surface proteins selected among previously characterized antigens, published in the literature (search terms - *Toxocara canis, recombinant proteins, membrane proteins, surface proteins, therapeutic proteins*). Sequences corresponding to membrane proteins (i.e. those known to be in direct contact with host immune cells and likely to mediate protective immunity) [22,25,27] were selected, particularly those with higher numbers of B cell epitopes. Nucleotide sequences corresponding to these proteins were analyzed by BLASTX (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify homologous sequences, by PSORT (http://www.psort.org/) to determine cellular localization, and by TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/) and SOSUI (http://bp.nuap.nagoya-u.ac.jp/sosui/) to identify signal peptide and transmembrane helices. ScanPROSITE (http://au.expasy.org/prosite/) was used to identify conserved domains and regions within the sequences. This analysis resulted in the selection of 4 genes rTcGPRC [Tcan_03886 (KHN73740.1)]; rTcCad [Tcan_06969 (KHN87575.1)]; rTcVcan [Tcan_14421 (KHN75413.1)] and rTcCyst [Tcan_18226 (KHN76703.1)] for expression of their codified proteins.

2.2. Identification and selection by proteomics analysis of potential immunoreactive targets

In addition to the *in silico* approach, a proteomics analysis of the excreted/secreted products (TES) of *T. canis* was done using a reverse phase nano-HPLC coupled to a nanoelectrospray and a Q Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific) to identify putative molecules for immunodiagnosis [16]. Subsequently, *in silico* analysis identified sequences that did not cross-react with other helminths and/or aeroallergens (*Ascaris lumbricoides, Trichuris trichiura, Ancylostoma* spp., *Schistosoma mansoni, Blomia tropicalis, Periplaneta americana, Blatella germanica and Dermatophagoides pteronyssinus*) yielding the proteins rTES26 (A0A0B2UWT5), rTES32 (O44927), rMUC-3 (Q9U9J2) and rCTL-4 (Q9XYV5). Immunoreactivity of these proteins was determined by ELISA and showed high sensitivity and specificity to detect specific antibodies in human sera known to be seropositive for *Toxocara* spp. (da Silva et al., manuscript in submission).

2.3. Expression of in silico selected recombinant antigens

Nucleotide sequences corresponding to the proteins selected *in silico* were synthesized commercially (GenScript Inc, Piscataway NJ, USA) and cloned into the expression vector pET-21a (+). Plasmids containing nucleotide sequences encoding rTcGPCR, rTcVcan, rTcCyst and rTcCad were transformed by thermal shock in BL21 (DE3) *E. coli* strain. Protein expression was done by cultivation in Luria Bertani (LB; Difco™, Le Pont de Claix 38800, France) at pH 7.4, and containing 100 μg/ml of ampicillin at 37 °C, to which 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (Thermo Scientific, Waltham, MA, USA) was added. After addition of IPTG inducer, cultures were stirred for 4 h until maximum expression of target proteins, except for rTcCyst that required 1 h of culture. Protein expression was confirmed in 12% polyacrylamide gels (SDS-PAGE), followed by Western blotting (WB), as described (Pinheiro et al., 2014). After confirmation of expression of recombinant proteins, corresponding bacterial pellets were treated with solutions for native and denaturing conditions to determine solubility. Cultures were centrifuged (10,000*g* × 30 min) at 4 °C and resulting pellets resuspended in lysis buffer (10 mM Na2HPO4, 10 mM NaH2PO4, 0.5 M NaCl) pH 7.4 and subjected to 5 cycles of 45 Hz sonic pulses for 1 min with 1-minute rest in ice between pulses, using a Q55/CL-188 Sonicator (Q Sonicators, Newtown, USA). After centrifugation (10,000*g* × 15 min); supernatants corresponding to the soluble fractions were collected and pellets resuspended again in denaturing lysis buffer (10 mM Na2HPO4, 10 mM NaH2PO4, 0.5 M NaCl, 8 M U) followed by further centrifugation and collection of supernatants corresponding to insoluble fractions. Super-

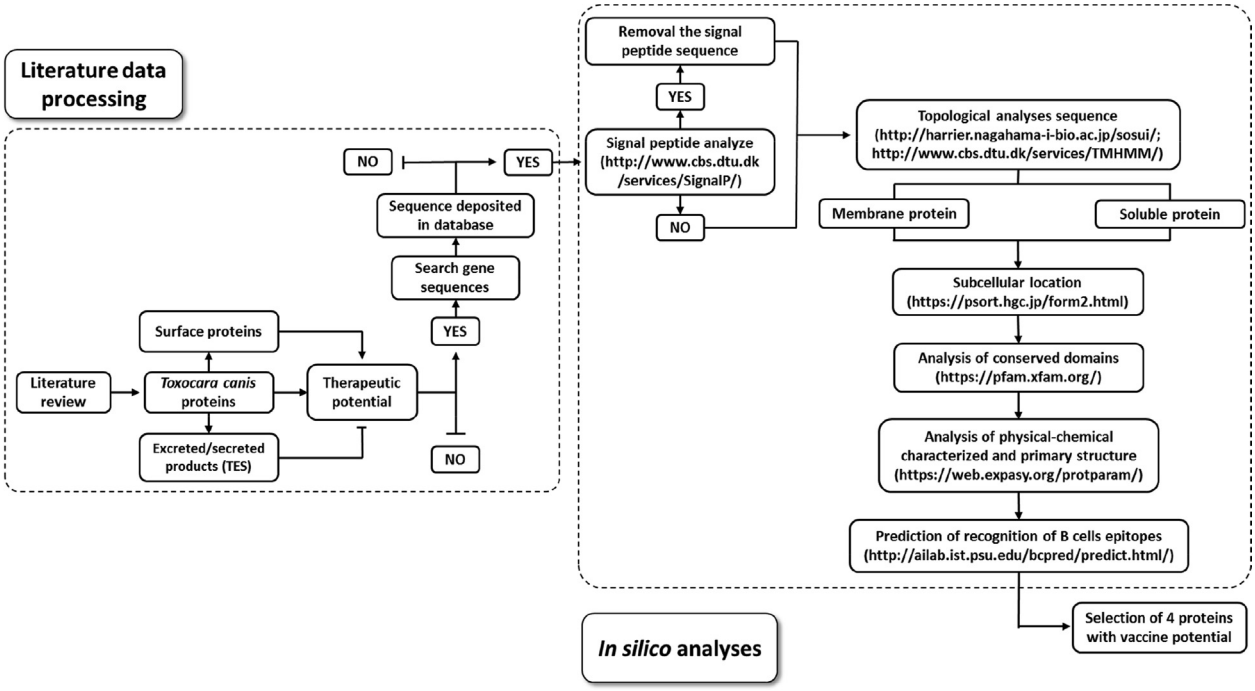


Fig. 1. Flow chart showing *in silico* analyses used for selection of vaccine candidate antigens by literature search and molecule characterization.

natants were analyzed by 12% polyacrylamide gel (SDS-Page). The rTcCad, rTcVcan and rTcCyst proteins were purified by automated affinity chromatography on AKTA Pure 25® from GE Healthcare (GE Healthcare, Bio-Sciences AB, Björkgatan 30 751 84 Uppsala, Sweden) with a HisTrap FF column (GE Healthcare, Bio-Sciences AB, Björkgatan 30 751 84 Uppsala, Sweden) using denaturing buffer (10 mm Na2HPO4, 10 mM NaH2PO4, 8 M urea and 500 mM imidazole). Dialysis was performed on a AKTA Pure 25 GE Healthcare using desalting column (GE Healthcare, Bio-Sciences AB, Björkgatan 30 751 84 Uppsala, Sweden) with 10 mM phosphate buffer. Endotoxin levels were measured, and depletion done when elevated. Dosage of recombinant proteins was performed with the Quibit protein assay dosing kit (Thermo Fisher Scientific, Eugene, Oregon, USA), and the recombinant proteins were stored at −70 °C until use.

2.4. Canine sera

Serum samples from 20 dogs from different areas of the metropolitan region of Salvador were obtained at the Teaching Hospital of Veterinary Medicine of the Federal University of Bahia. Dogs were examined clinically and blood obtained by venous puncture of the cephalic vein. Informed consent was obtained by the Municipal Sanitary Surveillance and the protocol approved by the Ethics Committee for Animal Studies of the Faculty of Veterinary Medicine of the Federal University of Bahia (CEUA n. 13/2017), following the International Guiding Principles for Biomedical Research Involving Animals.

2.5. Excretory/secretory T. canis larval antigens (TES)

The method of Savigny (1975) as modified (Alcântara-Neves et. al, 2008) was used to obtain excretory/secretory antigens of *T. canis* larvae (TES). Protein content was determined using the Lowry method (BioRad, Protein Assay, USA).

2.6. Pre-absorption of dog sera with somatic antigens of Ancylostoma caninum, Spirocerca lupi and Dirofilaria immitis

Because the development of a canine vaccine is our main objective, dog sera were pre-absorbed using antigens from helminth parasites relevant to these animals. *Ancylostoma caninum*, *S. lupi,* and *D. immitis* adult worms (obtained at the Teaching Hospital of Veterinary Medicine of the Federal University of Bahia and from Dr. Simone Campos, Veterinary Pathologist of Rio de Janeiro) were washed with PBS, triturated, subjected to thermal shock with liquid nitrogen and water, ground in a sonicator (model CL-188, QSonica Sonicators, Newtown, USA) with 70 Hz amplitude pulses in PBS containing PMSF (Sigma Aldrich, St Louis, MO, USA). The suspensions were centrifuged three times to remove cell debris and residual fats and the soluble fractions were aliquoted and stored at −70 °C. Protein content was determined by the Lowry method (BioRad, Protein Assay, USA). Pre-absorption of sera was done using a 1:5 dilution of canine sera, incubated with 4.0 mg/mL of somatic antigen from *A. caninum, S. lupi* and *D. immitis* in PBS containing 15% polyethylene glycol (PEG 15,000, Sigma Chemical Co., St. Louis, MO, USA) and 0.1% azide (Sigma Chemical Co., St. Louis, MO, USA). Sera were homogenized for 30 min and centrifuged at 5724*g* for 10 min at 4 °C, and supernatants were stored at −20 °C.

2.7. Immunoenzymatic assay for detecting anti-Toxocara spp. IgG in dog sera

Anti-*Toxocara* spp. IgG antibodies were detected using an indirect ELISA as previously described [28] in which TES products were used as antigens. In the present assay, we used 3 μg/mL of recombinant proteins (rTcVcan, rTcCyst, or rTcCad) or 3.5 μg/mL of TES for coating plates. Negative control was a serum pool from ten dogs born to dewormed mothers and kept in relative isolation and treated periodically with anthelmintic drugs (i.e. at least three times a year).

Table 1
Summary of immunization groups.

	Groups	Recombinant proteins	Adjuvant
First experiment	1	Negative control	No
	2	Negative control	Yes
	3	rTcCad	Yes
	4	rTcVcan	Yes
	5	rTcCyst	Yes
	6	rTcCad, rTcVcan and rTcCyst (Mix 1.0)	Yes
	7	rTcCad, rTcVcan and rTcCyst (Mix 1.1)	No
Second experiment	1	Negative control	No
	2	Negative control	Yes
	3	rTcCad	Yes
	4	rTES26	Yes
	5	rTES32	Yes
	6	rMUC-3	Yes
	7	rCTL-4	Yes
	8	rTES26, rTES32, rMUC-3 and rCTL-4 (Mix 2.0)	Yes

2.8. Immunization and challenge of mice with the recombinant proteins of Toxocara canis

For the development of the animal model, all procedures were done in accordance with national regulations on animal experimentation and welfare, and were authorized by the ethics commission for the use of animals (CEUA) of the Institute of Health Sciences (ICS)/UFBA (permit CEUA N° 137/2018).

The mice were fed with regular diet and water *at libitum*. Female C57BL/6, kept in sterile cages, were divided into 7 immunization groups of 7 mice per group for the first experiment: Group 1 - PBS; Group 2 – PBS with adjuvant; Group 3 –rTcCad; Group 4 – rTcVcan; Group 5 - rTcCyst; Group 6 - mixture of rTcCad, rTcVcan and rTcCyst with adjuvant (Mix 1.0); Group 7- mixture of rTcCad, rTcVcan and rTcCyst with PBS (Mix 1.1). In a second experiment, C57BL/6 females were divided into 8 groups of 7 mice/group for immunization with: Group 1 – PBS; Group 2 PBS with adjuvant; Group 3 – rTcCad (repeat of first experiment); Group 4 - rTES26 protein; Group 5 - rTES32 protein; Group 6 - rMUC-3; Group 7- rCTL-4 protein; Group 8 - mixture of rTES26, rTES32, rMUC-3 and rCTL-4 with adjuvant (Mix 2.0). The immunization protocol was as follows: mice were immunized subcutaneously with 25 μg of the recombinant proteins with either Freund's complete adjuvant (ACF, SIGMA-ALDRICH Co., St. Louis, MO, USA) or PBS control (Table 1) in a total volume of 100 μL/animal on day 0. Mix group immunizations used an equal concentration of individual recombinants to make a total concentration of 25 μg. On day 7th, mice were immunized with the same dose of recombinant proteins with Freund's incomplete adjuvant (AIF, SIGMA-ALDRICH Co., St. Louis, MO, USA) or PBS control. A third immunization was done on day 14 with the same solution used in the second immunization. After the final vaccination on day 14, all groups were challenged orally with 500 embryonated *T. canis* eggs in 100 μL saline, and the mice were kept for 30 days until sacrificed. The vaccination schedule is summarized in Table 1 and illustrated in Fig. 2.

Table 1. Summary of immunization groups.

2.9. Sample collection from mice

The mice were anesthetized by intraperitoneal injection of 250 μL of Xylazine and Ketamine at a ratio of 2:1 (40 mg/kg; Anasedan, São Paulo, Brazil); blood was collected from the brachial vein for antibody measurements, and after sacrifice, spleens were removed aseptically for splenocyte cultures and cytokine measurements. Liver, brain and muscle were removed and frozen for larval quantification.

2.10. Evaluation of murine antibody responses

Microwell plates were incubated with 3 μg of recombinant proteins (or 3 μg of each proteins for mixes) in carbonate/bicarbonate buffer (18 mM Na2CO3, 45 mM NaHCO3, pH 9.6) overnight at 4 °C. Plates were blocked with 0.05% Tween solution/10% FBS (Sigma-Aldrich, St. Louis, MO, USA) in PBS for 1 h at room temperature, and then sera were added at a dilution of 1/500. Biotin-conjugated detection antibodies against mouse IgG (BD Bioscience, San Diego, CA, USA) and IgG2a (BD Bioscience, San Diego, CA, USA) were added at dilutions of 1/1000 and 1/125, respectively. All incubations were done with PBS-Tween (0.05%)-SBF (2.5%) as diluent, at room temperature and plates were washed 5X with PBS-Tween (0.05%) between each step. Streptavidin-Peroxidase (BD Bioscience, San Diego, CA, USA) was added at 1/1000 for 30 min. Development was done with TMB and hydrogen peroxide (Sigma-Aldrich; St. Louis, MO, USA) and the plates were read at 450 nm. For IgE detection, microwell plates were coated with 3 μg/ml of recombinant proteins using the ELISA protocol described above except that sera were diluted 1/5 and the secondary antibody was a biotinylated mouse anti-mouse IgE secondary antibody (BD Bioscience, San Diego, CA, USA) diluted at 1/250.

2.11. Splenocytes cultures and quantification of cytokines

Mice spleens were macerated and suspended in RPMI 1640 supplemented with 10% inactivated fetal bovine serum 10 mM HEPES buffer, 0.2 mM L-glutamine and 5 mg gentamicin/mL (Sigma-Aldrich, St. Louis, MO, USA). Cells were stimulated with 5 μg/well of the recombinant proteins for 48 h in 96-well plates at 37 °C in 5% CO_2 at a concentration of 0.8–1.0 × 10^5 cells/mL.

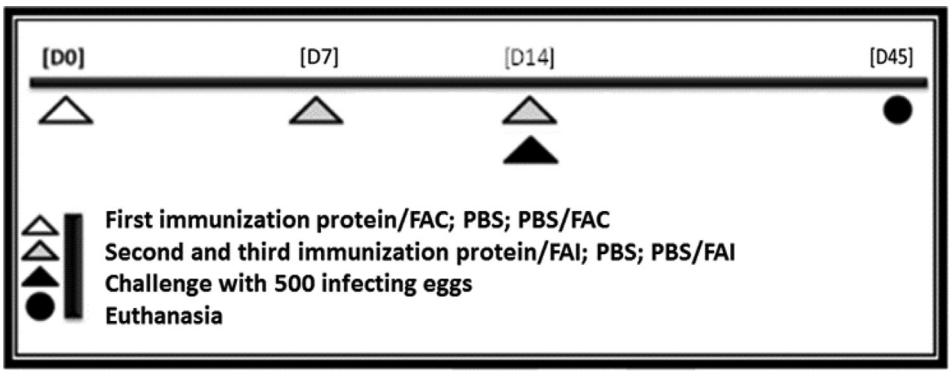


Fig. 2. Protocol for immunization of mice in murine model of toxocariasis. D – Day. FAC – Complete Freund's Adjuvant. FAI – Incomplete Freund's Adjuvant.

Cell-free supernatants were stored at −80 °C. IL-5, INF-ɣ and IL-10 in supernatants were quantified using commercial ELISA assays (OptEIA, BD Bioscience, San Diego, CA, USA)) following the manufacturer's instructions.

2.12. Harvesting of larvae from tissues of Toxocara canis challenged mice

Cryopreserved livers, brains and muscles were digested following established protocols [29,30]. Tissues were digested in a solution of pepsin 0.05% (Sigma-Aldrich, Saint Louis, MA, USA) in 0.063 M HCl (pH 1.5) with an enzymatic activity of 5000 units FIP/g (Fungi Lipase-International FIP Standard) [31] at 37 °C overnight. After digestion, lysates were transferred to polystyrene screens with pore size <20 μm, and then washed in sterile distilled H_2O thrice. The suspensions were transferred to 15 ml Falcon sterile tubes for centrifugation at 3000*g* for 5 min at 4 °C. The resulting pellets were resuspended and aliquots of 100 μL were collected for analysis by microscopy for counting of larvae. With this protocol, larvae are dead but intact, thus allowing quantitation.

2.13. Statistical analysis

The Kolmogorav Smirnov and Shapiro Wilk tests were used to evaluate normality of immunological data. Where appropriate, Wilcoxon signed-rank test or paired t-tests were used to compare paired groups of continuous measurements. ANOVA was used to evaluate heterogeneity across independent groups with post hoc Dunn's multiple comparison tests done for inter-group comparisons (i.e. vaccinated versus adjuvant control groups) with statistical significance inferred by $P < 0.05$. GraphPad Prism 7.0 was used.

3. Results

3.1. Selection of vaccine candidates by in silico analysis and production of recombinant proteins

Of thirty sequences corresponding to membrane proteins, four that fulfilled the criterion for number of B cell epitopes were selected: "protein homologue to receptor coupled to protein G (TcGPCR [Tcan_03886 (KHN73740.1)]), protein homologue to channel gene of potassium (TcVcan [Tcan_14421 (KHN75413.1)]), "protein homologue to protein gene cystatin (TcCyst [Tcan_18226 (KHN76703.1)]) and "protein homologue to cadherin (TcCad [Tcan_06969 (KHN87575.1)]). The genes encoding the proteins TcGPCR and TcVcan were modified excluding the sequences of signal peptide and transmembrane helices, and the genes encoding the proteins TcCyst and TcCad were modified excluding the signal peptide sequence. All gene constructions encoding these four modified proteins were synthesized, cloned, and expressed (Fig. 1 supplementary) except rTcGPCR that could not be expressed. The rTcCad, rTcVcan and rTcCyst were expressed in the insoluble fraction and purified under denaturing conditions (Fig. 1 supplementary) with yields of 250 to 562 μg/mL purified protein.

3.2. Recognition of in silico selected recombinant proteins by sera from infected dogs

Fig. 3 shows IgG reactivity in seropositive dog sera (i.e. anti-TES IgG positive) against TES antigen and recombinant proteins before and after absorption against somatic antigens of *A. caninum, S. lupi* and *D. immitis*. Canine sera reacted weakly against recombinant proteins. Absorption significantly reduced optical densities only for TES antigens, indicating cross-reactivity for TES antigens but minimal cross-reactivity for recombinant proteins.

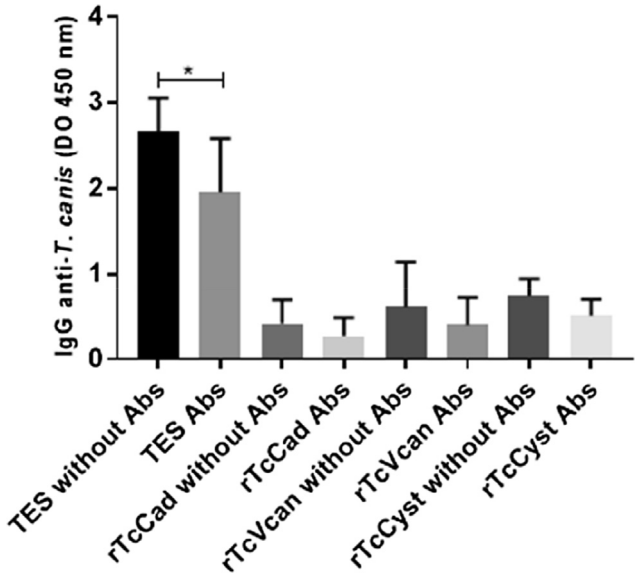


Fig. 3. IgG reactivity by indirect ELISA against excreted/secreted antigens of *T. canis* (TES) and *in silico*-identified membrane-derived recombinant proteins rTcCad, rTcVcan and rTcCyst, before (without) and after absorption to somatic antigens of *Ancylostoma caninum*, *Spirocerca lupi* and *Dirofilaria immitis*. Columns and bars represent means and standard errors, respectively, of ODs from 20 dog sera. Analyses by Wilcoxon matched pairs signed rank test. *p < 0.005 for difference between paired absorbed (Abs) and non-absorbed (without Abs) sera *p < 0.005.

3.3. Antibody responses after immunization of mice with recombinant proteins

Fig. 4 shows IgG, IgG1, IgG2a and IgE-specific responses in mice vaccinated with seven recombinants proteins. Results of two experiments are shown (Fig. 4). In the first experiment (Fig. 4, A1-D1), mice were vaccinated with *in silico*-identified recombinant proteins and protein mixes with and without adjuvant (rTcCad, rTcVcan, rTcCyst, Mix 1.0 [rTcCad, rTcVcan and rTcCyst with adjuvant], and Mix 1.1 [rTcCad, rTcVcan and rTcCyst without adjuvant]). Elevated levels of specific antibodies to the following recombinant protein were observed compared to controls (adjuvant only): all antibodies for rTcCad and rTcVcan; IgG1 for rTcCyst; IgG1, IgG2a, and IgE for Mix 1.0; and IgG and IgG1for Mix 1.1. In the second experiment (Fig. 4, A2-D2), a repeat vaccination was done with rTcCad and mice were vaccinated with candidate recombinants proteins identified by proteomics analyses (rTES26, rTES32, rMUC-3, rCTL-4 and Mix 2.0 [rTES26, rTES32, rMUC-3, and rCTL-4]). Significantly, elevated levels of specific antibodies to the following recombinant proteins were observed compared to adjuvant only controls: all antibodies for rTcCad, rTES32, rMUC-3, and Mix 2.0; and IgG, IgG2a, and IgE for rCTL-4. In both experiments there were trends of greater IgG1 responses to *in silico*-identified recombinant proteins and for one of these, rTcCad stronger IgG2a responses. Recombinant protein mixes did not induce stronger responses than those of the individual proteins.

3.4. Cytokine production by splenocytes from mice vaccinated with recombinant proteins

Fig. 5 shows levels of IL-5, IFN-γ, and IL-10 produced by splenocytes harvested from mice vaccinated with recombinant proteins and protein mixes for the two experiments. Significantly elevated levels of IL-5 were observed for all *in silico* identified recombinant proteins (rTcCad, rTcVcan, rTcCyst) and mixes (Fig. 5, A1), but for none of the proteomics identified recombinant proteins (rTES26, rTES32, rMUC-3, rCTL-4) or proteins mixes (Fig. 5, A2). Immuniza-

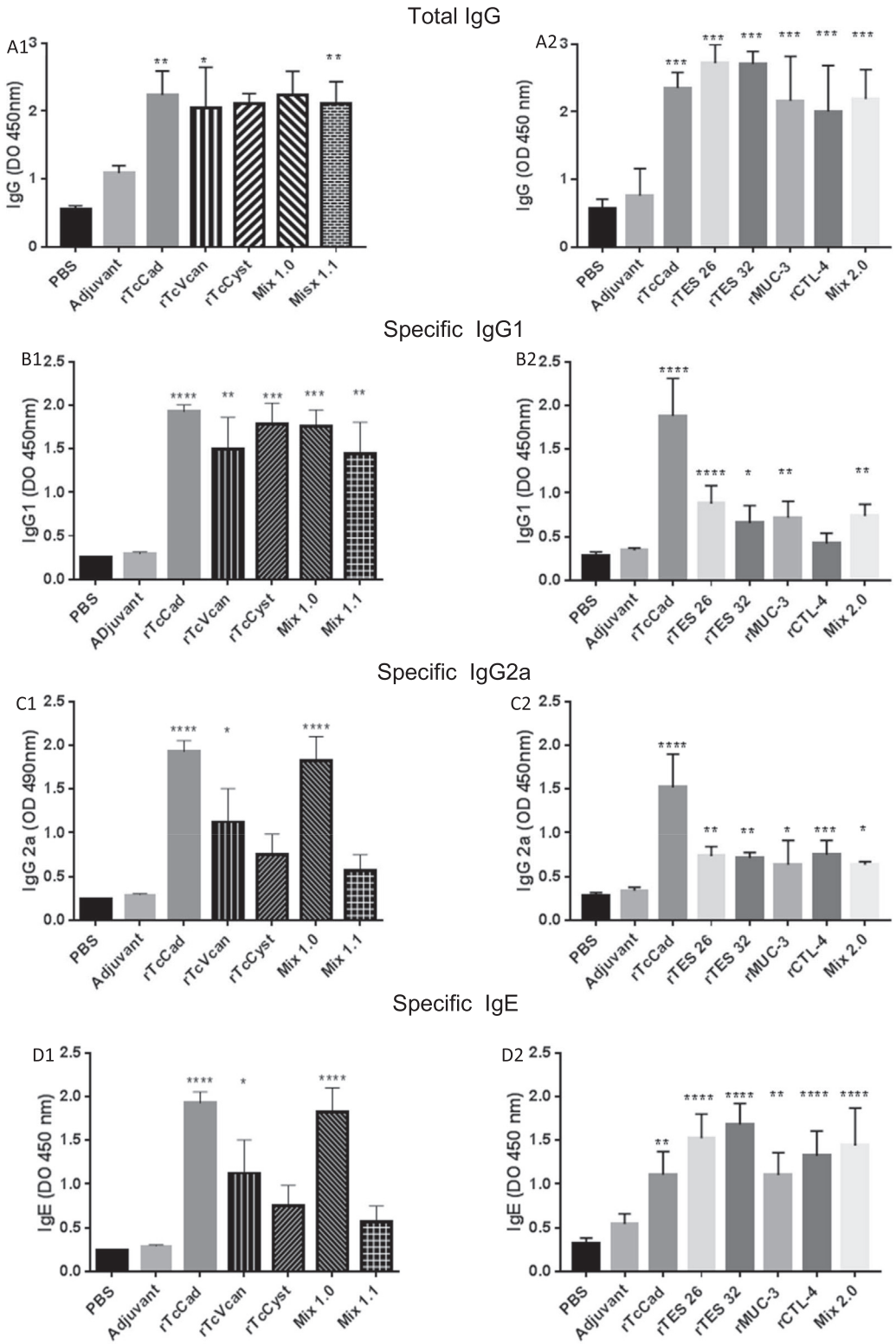


Fig. 4. Serum responses of IgG (A1,A2), IgG1 (B1, B2), IgG2a (C1, C2) and specific IgE (D1, D2) in mice vaccinated with rTcCad, rTcVcan, rTcCyst, rTES26, rTES32, rMUC-3, rCTL-4 and Mix 1.0, Mix 1.1. and Mix 2.0. Asterisks indicate statistically significant differences between unvaccinated (adjuvant control) and individual vaccination groups in Dunn's post hoc test: **** $p < 0.0001$; *** $p < 0.001$, ** $p < 0.01$ and * $p < 0.05$.

tions with all individual recombinant proteins and Mix 2.0 were associated with increased production of IFN-γ (Fig. 5, B1 and B2). All recombinant proteins and mixes were associated with elevated levels of IL-10 (Fig. 5, C1 and C2). For *in silico*-identified recombinant proteins: greater IL-10 production was observed for mice immunized with rTcCyst ($p = 0.0001$) compared to rTcCad (p 0.008) and rTcVcan ($p = 0.007$); recombinant protein mixes did not induce stronger cytokine responses than those of the individual proteins; and IFN-γ responses to individual recombinants

were greater compared to those for adjuvant-Mix 1.0 (rTcCad, p = 0.03; TcVcan, p = 0.003; and rTcCyst, p = 0.008).

3.5. Evaluation of the effects of recombinant proteins on larval survival after challenge

Fig. 6 shows the number of larvae harvested from muscle, liver, and brain following infectious challenge of mice vaccinated with recombinant proteins and protein mixes. For larvae in muscle (Fig. 6, A1 and A2), significantly reduced counts were observed for all 7 mice vaccinated with recombinant proteins compared to adjuvant alone, although reductions were greater for the 3 *in silico*-identified molecules (rTcCad 60.4% reduction, p = 0.001; rTcVcan 58.1%, p = 0.002; and rTcCyst 50.9%, p = 0.008) than proteomics-identified molecules (Fig. 6, A2: rTcCad vs. rTES26, p = 0.0008; vs. rTES32, p = 0.04; vs. rMUC-3, p = 0.002; vs. rCTL-4, p = 0.0002). In the liver, significantly reduced counts were observed only for rTcCad and rTcVcan (rTcCad 52.0% and rTcVcan 53.2%, both with p < 0.005) vaccinated mice. Larval counts in the liver were significantly reduced to compared to rTcCad compared to proteomics-identified recombinant proteins (Fig. 6, B2: rTcCad vs. rTES26, p = 0.002; vs. rTES32, p = 0.0001; vs. MUC-3, p = 0.02; vs. rCTL-4, p = 0.03; vs. Mix 2.0, p = 0.02). Larval yields in the brain did not differ significantly from adjuvant control for any of the 7 recombinant proteins (Fig. 6, C1 and C2). Reductions in total larval counts from all organs were greater for rTcCad

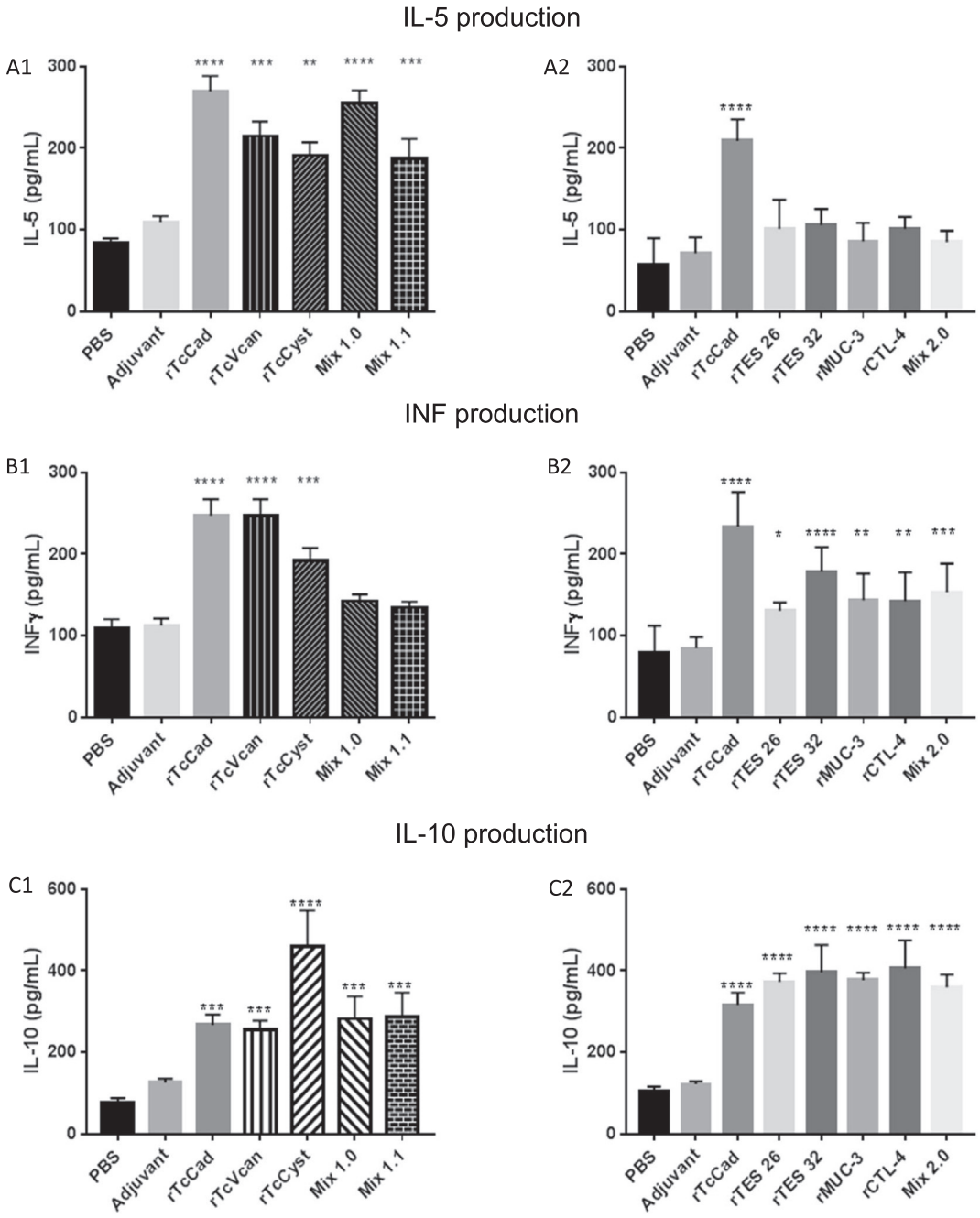


Fig. 5. Cytokines produced by cultivated spleen monocytes from mice vaccinated with rTcCad, rTcVcan, rTcCyst, rTES26, rTES32, rMUC-3, rCTL-4, Mix1.0, Mix 1.1 and Mix 2.0. Results from the two experiments are shown for IL-5 (A1, A2), INF-γ (B1, B2) and IL-10 (C1, C2). Asterisks indicate statistically significant differences between unvaccinated (adjuvant control) and individual vaccination groups in Dunn's post hoc test: **** p < 0.0001; *** p < 0.001, ** p < 0.01 and * p < 0.05.

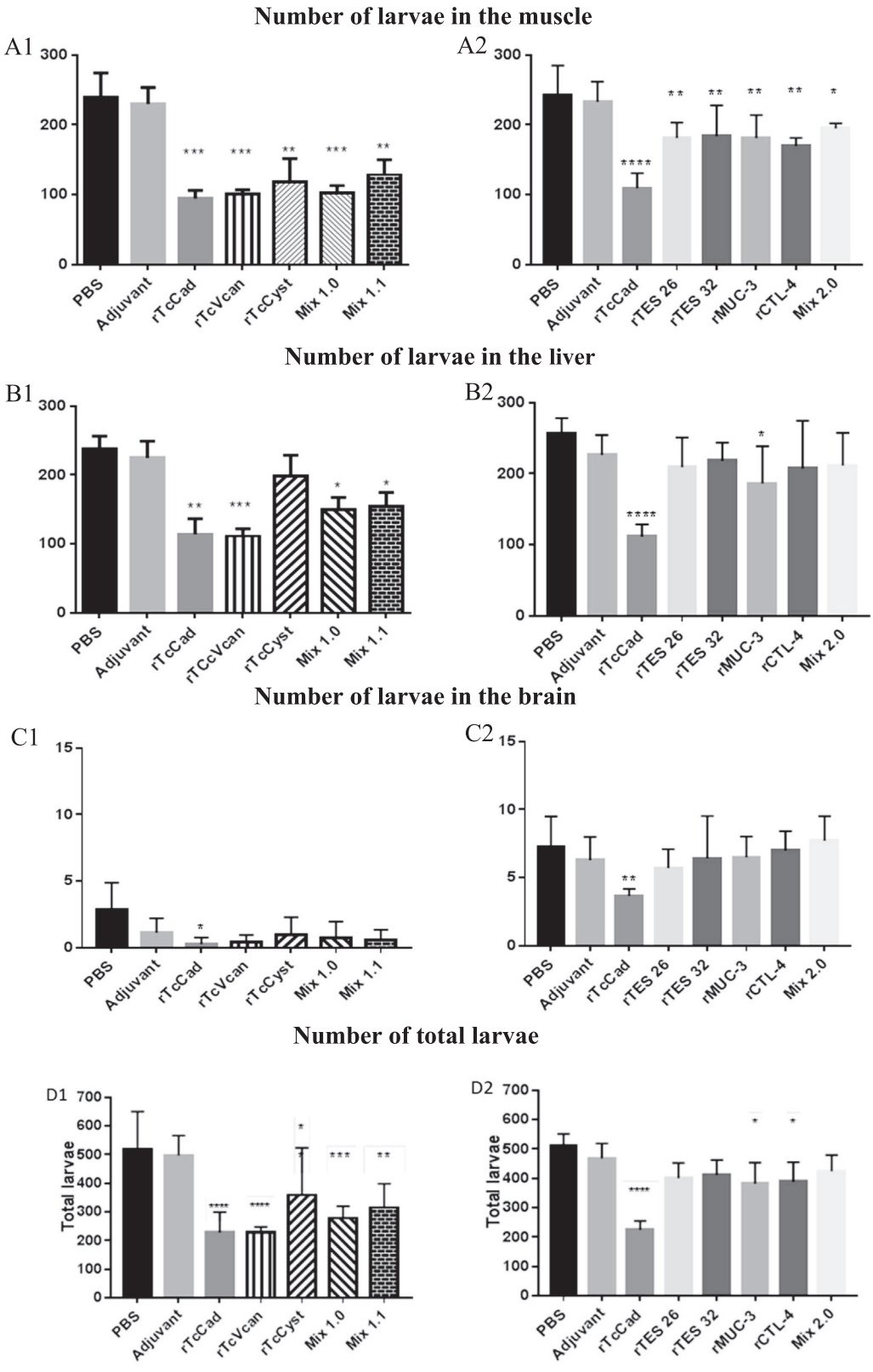


Fig. 6. Number of *T. canis* larvae recovered from muscle (A1), liver (B1), brain (C1) and all tissues (D1) from mice vaccinated with rTcCad, rTcVcan, rTcCyst, Mix 1.0 and Mix 1.1 Results from the two experiments are shown for numbers of larvae harvested from muscle (A1 & A2), liver (B1 & B2), brain (C1 & C2), and total number of larvae (D1 & D2). After vaccination, mice were challenged with 500 *T. canis* embryonated eggs. Asterisks indicate statistically significant differences between unvaccinated adjuvant controls and vaccination groups in Dunn's post hoc test: **** $p < 0.0001$; *** $p < 0.001$, ** $p < 0.01$ and *$p < 0.05$.

(53.4%) and rTcVcan (53.9%) than for the other recombinant proteins or mixes (rTcCyst 31.3%, Mix 1.0 45.2%, Mix 1.1 41.3%; rTES26 15.7%, rTES32 16.7%, rMUC-3 20.4%, rCTL4 18.7%, and Mix 2.0 15.4%). Significantly reduced larval counts (Fig. 6, D1 and D2) were observed for 5 recombinant proteins compared to adjuvant alone (rTcCad 54.3% reduction, $p < 0.0001$; rTcVcan 53.9%, $p < 0.0001$; rMUC-3 18.4%, $p = 0.02$; and rCTL-4 16.7%, $p = 0.04$) although reductions appeared to be greater for 2 *in silico*-identified recombinant proteins (rTcCad and rTcVcan). Reductions in larval counts in mice immunized with rTcCad were greater than for the proteomics-identified recombinants (Fig. 6 D2: rTcCad vs. rTES26, $p = <0.0001$; vs. rTES32, $p = 0.0008$; vs. rMUC-3, $p = 0.0011$; vs. rCTL-4, $p = 0.0008$).

4. Discussion

Technological developments now allow the identification of "hidden" antigens (i.e. those to which the host immune response is not normally exposed) that are potentially useful targets for interference with the nutritional and metabolic functions of parasites and cause their expulsion or death [19,21,23]. These new technologies, including *in silico* and proteomics analyses used in the present study, allow the identification and characterization of candidate molecules for vaccines and diagnostics according to their locations, characteristics and potential functions in the parasite [32–34].

In the present study, we identified 4 proteins from the proteome of *T. canis* [16] shown previously to have potential as novel diagnostic antigens (da Silva et al., manuscript submitted). Using *in silico* analyses, we identified 30 protein sequences representing surface proteins, of which we selected 4 based on immunogenic potential (i.e. likely number of B cell epitopes). Three of these 4 sequences (rTcVcan, rTcCyst and rTcCad) were expressed and purified and appeared to have minimal IgG cross-reactivity with other canine helminths (*A. caninum, S. lupi* and *D. immitis)* in absorption studies using *Toxocara*-seropositive canine sera.

We tested seven recombinant proteins for immunogenicity and protection in a murine model of *T. canis* infection in two experiments. In the first, mice were immunized with the membrane proteins (rTcCad, rTcVcan and rTcCyst). In the second experiment, mice were vaccinated with those proteins from TES identified by proteomics (rTES26, rTES32, rMUC-3 and rCTL-4) concomitantly with rTcCad, because of the promising results obtained for this protein in the first experiment. Immunizations with rTcCad protein stimulated a mixed antibody response with greater production of IgG1, IgG2a, and IgE compared to the other membrane proteins, rTcVcan and rTcCyst. The same immune response for rTcCad was observed in the 2nd experiment, which induced greater production of IgG1 and IgG2a compared to the proteomics identified recombinant proteins (rTES26, rTES32, rMUC-3, rCTL-4). However, these proteins induced similar or greater IgE production compared to rTcCad. In both experiments we immunized animals using complete and incomplete Freund adjuvant as recommended previously for murine models of vaccine candidates against helminth infections [35–37]. Having identified the best vaccine candidates in the present study, we are planning to select the best adjuvant to be used in combination with these candidate antigens using the same murine model.

Strong antibody-mediated immune responses are characteristic of helminth infections, and it has been suggested that a Th1/Th2 mixed response may be more advantageous than a polarized response for protective immunity, as shown by vaccination studies using recombinant molecules derived from *Schistosoma mansoni, Ascaris suum* and *Ancylostoma caninum* [38–42].

In the present study, reduction in larval burdens following infectious challenge with *T. canis* eggs might be mediated partly by binding of anti-rTcCad and anti-rTcVcan IgG and IgE to the cuticle of *T. canis* larvae where these antigens are present, inducing larval killing through antibody-dependent cellular cytotoxicity (ADCC) reactions in the presence of activated macrophages or/and eosinophils [43–46].

Elevated IL-5 induced by these membrane-associated proteins (rTcCad and anti-rTcVcan) could enhance eosinophilic activation at the sites of larval infection in the tissues and larval killing through the release of toxic eosinophilic granules. Elevated INF-γ may enhance macrophage activation and the induction of nitric oxide synthase (iNOS) which is deleterious to metabolic processes within the parasite [47].

Different studies have investigated potential vaccine candidates against *T. canis*. A previous murine model of vaccination against *T. canis* showed moderate protection (40.5%) induced by TES in IFA (40.5%) following challenge with 1000 embryonated eggs of *T. canis* [36]. Other studies showed approximately 30% reduction after challenge with *T. canis* in mice immunized with *T. canis* somatic antigens or TES in CFA [35,48].

Although IgG is the dominant antibody isotype present in murine toxocariasis, there is evidence for the preferential production of the IgG1 isotype [49,50], as observed in the present study, following vaccination with membrane proteins. The IgG1 subclass in mice is considered to be an anaphylactic (eosinophil-dependent) antibody, capable of inducing histamine release by mast cells [50].

The stimulation of IgE and IgG1 antibodies in mice after infection with *T. canis* is consistent with previous studies [29,49–51]. However, a TES-immunization model in mice showed that one to two immunizations with TES antigens (in IFA) [52] induced much lower levels of IgG1 and IgG2 antibodies in the context of reduced larval counts following infectious challenge with *T. canis* eggs. Such observations could indicate that protection is associated with a shift from a polarized Th2 to a mixed Th1/Th2 antibody response.

Immunoglobulin isotype-specific responses observed in mice immunized with rTcCad and rTcVcan were consistent with the profiles of cytokine produced by splenocyte cultures (i.e. enhanced production of both Th1 [IFN-γ] and Th2 [IL-5] cytokines) indicating that these recombinant molecules induce a mixed Th1/Th2 immune response. On the other hand, rTcCyst induced elevated levels of immune regulatory IL-10 compared to rTcCad and rTcVcan, representing a stronger immunomodulatory response that may be undesirable in a candidate vaccine. A previous study using *A. suum* recombinant proteins showed protection against challenge with *A. suum* eggs to be associated with a mixed Th1/Th2 response [38]. Increased production of IL-5 by splenocytes from mice immunized with rTcCad and rTcVcan could have increased numbers of activated eosinophils in Peyer's patches of vaccinated animals, thus reducing the numbers of larvae crossing the intestinal mucosa to reach the liver via the portal vein [53–55].

Cytokine production by splenocytes from mice vaccinated with the proteomics-identified recombinant proteins (rTcCyst, rTES26, rTES32, rMUC-3 and rCTL-4) was characterized by a mixed IFN-γ/IL-10 response. The IL-10 response might explain increased larval survival following infectious challenge (as for the membrane protein rTcCyst), because of the role of IL-10 in the regulation of inflammatory responses (both Th1 and Th2-mediated) [56–58]. A mixed IL-5/IL-10 response, in the case of rTcCyst, has been referred to as a modified Th2 response [59–61].

We measured the survival of *T. canis* larvae in tissues of vaccinated mice challenged with *T. canis* eggs. Immunizations with the membrane-derived recombinant proteins, rTcCad and rTcVcan, caused the greatest reductions in larval counts post-challenge (approximately 55%). In a previous murine model of *T. canis,* TES products induced 40.5% protection while soluble somatic antigens

of L3 larvae induced 30% protection to infectious challenge [36]. The findings of our study, therefore, may support a strategy of identifying 'hidden' antigens as potential vaccine candidates [19,21,62]. Although TES-derived recombinant proteins were immunoreactive, they were less effective in reducing larval numbers following challenge, and may be more likely to induce host tolerance than protective responses [63,64].

In conclusion the development of a novel recombinant molecule-based vaccine represents a potential strategy for the control of *Toxocara* spp. infections in dogs and cats. Currently control of toxocariasis is based on periodic treatments of animals with incomplete coverage and the risk of development of drug resistance. To our knowledge, this is the first report in which a reverse vaccinology strategy, using *in silico* and proteomics analyses, has been used to identify new vaccine candidates for *Toxocara* spp. infections. Using these methods, we identified 7 immunogenic recombinant proteins of which two, rTcCad and rTcVcan, derived from the parasite membrane, induced a mixed Th1/Th2 immune response and reduced significantly (by approximately 53%) larval invasion of the tissues, after infectious challenge. The use of a murine model to identify candidate antigens is advantageous as it allows rapid prioritization of antigens through characterization of the immune response induced and evaluation of protection against infectious challenge. Once evaluated in the murine model, future studies are needed to optimize immunization protocols for the definitive hosts and define the mechanisms by which these recombinant proteins mediate protection.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors are grateful to FAPESB (Research's Foundation of the State of Bahia; grant APP0099/2016) and CNPq/INCT-Tera Nano for supporting this work. LFSG is receptor of the scholarship of FAPESB modality – PhD student (n.° BOL2255/2016). LACP and NMAN are receptors of CNPq scholarships for researchers.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vaccine.2020.04.072.

References

[1] Despommier D. Toxocariasis: clinical aspects, epidemiology, medical ecology, and molecular aspects. Clin Microbiol Rev 2003;16:265–72. https://doi.org/10.1128/CMR.16.2.265.
[2] Overgaauw P. Aspects of toxocara epidemiology: Human toxocarosis. Crit Rev Microbiol 1997;23:215–31. https://doi.org/10.3109/10408419709115137.
[3] Overgaauw P, van Knapen F. Veterinary and public health aspects of Toxocara spp. Vet Parasitol 2013;193:398–403. https://doi.org/10.1016/j.vetpar.2012.12.035.
[4] Dubná S, Langrová I, Nápravník J, Jankovská I, Vadlejch J, Pekár S, et al. The prevalence of intestinal parasites in dogs from Prague, rural areas, and shelters of the Czech Republic. Vet Parasitol 2007;145:120–8. https://doi.org/10.1016/j.vetpar.2006.11.006.
[5] Fok E, Szatmári V, Busák K, Rozgonyi F. Prevalence of intestinal parasites in dogs in some urban and rural areas of hungary. Vet Q 2001;23:96–8. https://doi.org/10.1080/01652176.2001.9695091.
[6] Habluetzel A, Traldi G, Ruggieri S, Attili AR, Scuppa P, Marchetti R, et al. An estimation of Toxocara canis prevalence in dogs, environmental egg contamination and risk of human infection in the Marche region of Italy. Vet Parasitol 2003;113:243–52. https://doi.org/10.1016/S0304-4017(03)00082-7.
[7] Lee ACY, Schantz PM, Kazacos KR, Montgomery SP, Bowman DD. Epidemiologic and zoonotic aspects of ascarid infections in dogs and cats. Trends Parasitol 2010;26:155–61. https://doi.org/10.1016/j.pt.2010.01.002.
[8] Overgaauw P. Prevalence of intestinal nematodes of dogs and cats in the Netherlands. Vet Q 1997;19:14–7. https://doi.org/10.1080/01652176.1997.9694730.
[9] Agudelo C, Villareal E, Cáceres E, López C, Eljach J, Ramírez N, et al. Human and dogs Toxocara canis infection in a poor neighborhood in Bogota. Mem Inst Oswaldo Cruz 1990;85:75–8. https://doi.org/10.1590/S0074-02761990000100012.
[10] Regis SCS, Mendonça LR, Silva Ndos S, Dattoli VCC, Alcântara-Neves NM, Barrouin-Melo SM. Seroprevalence and risk factors for canine toxocariasis by detection of specific IgG as a marker of infection in dogs from Salvador, Brazil. Acta Trop 2011;120:46–51. https://doi.org/10.1016/j.actatropica.2011.05.011.
[11] Katagiri S, Oliveira-Sequeira TCG. Prevalence of dog intestinal parasites and risk perception of zoonotic infection by dog owners in S??o Paulo State, Brazil. Zoonoses Public Health 2008;55:406–13. https://doi.org/10.1111/j.1863-2378.2008.01163.x.
[12] Gingrich EN, Scorza AV, Clifford EL, Olea-Popelka FJ, Lappin MR. Intestinal parasites of dogs on the Galapagos Islands. Vet Parasitol 2010;169:404–7. https://doi.org/10.1016/j.vetpar.2010.01.018.
[13] Maizels R, Savigny DH, Ogilvie BM. Characterization of surface and excretory-secretory antigens of Toxocara canis infective larvae. Parasite Immunol 1984:23–37.
[14] Holland CV, Smith H. Toxocara The Enigmatic Parasite; 2005.
[15] Zhu X, Korhonen PK, Cai H, Young ND, Nejsum P, Von Samson-himmelstjerna G, et al. Genetic blueprint of the zoonotic pathogen Toxocara canis. Nat Commun 2015;2:1–8. https://doi.org/10.1038/ncomms7145.
[16] da Silva MB, Urrego AJR, Oviedo Y, Cooper PJ, Pacheco LGC, Pinheiro CS, et al. The somatic proteins of Toxocara canis larvae and excretory-secretory products revealed by proteomics. Vet Parasitol 2018;259:25–34. https://doi.org/10.1016/j.vetpar.2018.06.015.
[17] Amerasinghe PH, Rajapakse RPVJ, Lloyd S, Fernando ST. Antigen-induced protection against infection withToxocara vitulorum larvae in mice. Parasitol Res 1992:643–7.
[18] Maizels R, Tetteh KKA, Loukas A. Toxocara canis: Genes expressed by the arrested infective larval stage of a parasitic nematode. Int J Parasitol 2000;30:495–508. https://doi.org/10.1016/S0020-7519(00)00022-9.
[19] Munn EA. Rational design of nematode vaccines: hidden antigens. Int J Parasitol 1997;27:359–66. https://doi.org/10.1016/S0020-7519(97)00003-9.
[20] Hotez PJ. Neglected infections of poverty in the United States of America. PLoS Negl Trop Dis 2008;2. https://doi.org/10.1371/journal.pntd.0000256.
[21] Knox DP, Redmond DL, Newlands GF, Skuce PJ, Pettit D, Smith WD. The nature and prospects for gut membrane proteins as vaccine candidates for Haemonchus contortus and other ruminant trichostrongyloids. Int J Parasitol 2003;33:1129–37. https://doi.org/10.1016/S0020-7519(03)00167-X.
[22] Andrews SJ, Rolph TP, Munn EA, Taylor MA. Duration of protective immunity against ovine haemonchosis following vaccination with the nematode gut membrane antigen H11. Res Vet Sci 1997;62:223–7. https://doi.org/10.1016/S0034-5288(97)90194-6.
[23] Pearson MS, Ranjit N, Loukas A. Blunting the knife: Development of vaccines targeting digestive proteases of blood-feeding helminth parasites. Biol Chem 2010;391:901–11. https://doi.org/10.1515/BC.2010.074.
[24] Gauci CG, Verástegui MR, Gilman RH, Lightowlers MW. Taenia solium and Taenia ovis: Stage-specific expression of the vaccine antigen genes, TSOL18, TSOL16, and homologues, in oncospheres. Exp Parasitol 2006;113:272–5. https://doi.org/10.1016/j.exppara.2006.01.006.
[25] Lightowlers MW, Rolfe R, Gauci CG. Taenia saginata: Vaccination against cysticercosis in cattle with recombinant oncosphere antigens. Exp Parasitol 1996;84:330–8. https://doi.org/10.1006/expr.1996.0121.
[26] Woollard DJ, Gauci CG, Heath DD, Lightowlers MW. Epitope specificities and antibody responses to the EG95 hydatid vaccine. Parasite Immunol 1998;20:535–40. https://doi.org/10.1046/j.1365-3024.1998.00176.x.
[27] Jasmer D, Perryman L, Conder G, Crow S, McGuire T. Protective immunity to Haemonchus contortus induced by immunoaffinity isolated antigens that share a phylogenetically conserved carbohydrate gut surface epitope. J Immunol 1993;151:5450–60.
[28] Savigny DHDE, Voller A, Woodruff AW. Toxocariasis: serological diagnosis by enzyme immunoassay. J Clin Pathol 1979:284–8.
[29] Concepcion JE, Barriga OO. Transfer of infection-induced immune protection to Toxocara canis in a mouse model. Vet Immunol Immunopathol 1985;9:371–82.
[30] Barriga OO, Myser WC. Effects of irradiation on the biology of the infective larvae of Toxocara canis in the mouse. J Parasitol 1987;73:89. https://doi.org/10.2307/3282349.
[31] Llarena-Reino M, Piñeiro C, Antonio J, Outeriño L, Vello C, González ÁF, et al. Optimization of the pepsin digestion method for anisakids inspection in the fishing industry. Vet Parasitol 2012;191:276–83. https://doi.org/10.1016/j.vetpar.2012.09.015.
[32] Aebersold R, Mann M. Mass spectrometry-based proteomics. Nat Publ Gr 2003;422.
[33] Phizicky E, Bastiaens PIH, Zhu H, Snyder M, Fields S. Protein analysis on a proteomic scale. Nature 2003;422:208–15. https://doi.org/10.1038/nature01512.
[34] Clauser KR, Baker P, Burlingame AL. Role of accurate mass measurement (±10 ppm) in protein identification strategies employing MS or MS/MS and database searching. Anal Chem 1999;71:2871–82. https://doi.org/10.1021/ac9810516.

[35] Sugane K, Kusama Y, Takamoto M, Tominaga A, Takatsu K. Eosinophilia, IL-5 level and recovery of larvae in IL-5 transgenic mice infected with Toxocara canis. J Helminthol 1996;70:153. https://doi.org/10.1017/s0022149x00015315.
[36] Dvoroznakova E, Borošková Z, Tomašovičová O. Immune responses in mice immunized with Toxocara canis antigens. Helminthologia 2002;39:59–66.
[37] Sansri V, Meemon K, Changklungmoa N, Kueakhai P, Chantree P, Chaichanasak P, et al. Protection against Fasciola gigantica infection in mice by vaccination with recombinant juvenile-specific cathepsin L. Vaccine 2015;33:1596–601. https://doi.org/10.1016/j.vaccine.2015.02.010.
[38] Islam MK, Miyoshi T, Tsuji N. Vaccination with recombinant Ascaris suum 24-kilodalton antigen induces a Th1/Th2-mixed type immune response and confers high levels of protection against challenged Ascaris suum lung-stage infection in BALB/c mice. Int J Parasitol 2005;35:1023–30. https://doi.org/10.1016/j.ijpara.2005.03.019.
[39] Loukas A, Bethony JM, Mendez S, Fujiwara RT, Narsa Goud G, Ranjit N, et al. Vaccination with recombinant aspartic hemoglobinase reduces parasite load and blood loss after hookworm infection in dogs. PLoS Med 2005;2:. https://doi.org/10.1371/journal.pmed.0020295e295.
[40] Tsuji N, Suzuki K, Kasuga-Aoki H, Matsumoto Y, Arakawa T, Ishiwata K, et al. Intranasal immunization with recombinant Ascaris suum 14-kilodalton antigen coupled with cholera toxin B subunit induces protective immunity to A. suum infection in mice. Infect Immun 2002;69:7285–92. https://doi.org/10.1128/iai.69.12.7285-7292.2001.
[41] Tsuji N, Suzuki K, Kasuga-Aoki H, Isobe T, Arakawa T, Matsumoto Y. Mice intranasally immunized with a recombinant 16-kilodalton antigen from roundworm ascaris parasites are protected against larval migration of Ascaris suum. Infect Immun 2003;71:5314–23. https://doi.org/10.1128/iai.71.9.5314-5323.2003.
[42] Tendler M, Brito CA, Vilar MM, Serra-Freire N, Diogo CM, Almeida MS, et al. A Schistosoma mansoni fatty acid-binding protein, Sm14, is the potential basis of a dual-purpose anti-helminth vaccine. Proc Natl Acad Sci 1996;93:269–73. https://doi.org/10.1073/pnas.93.1.269.
[43] Allen JE, Maizels RM. Diversity and dialogue in immunity to helminths. Nat Rev Immunol 2011;11:375–88. https://doi.org/10.1038/nri2992.
[44] Makepeace BL, Martin C, Turner JD, Specht S. Granulocytes in helminth infection - who is calling the shots?. Curr Med Chem 2012;19:1567–86. https://doi.org/10.2174/092986712799828337.
[45] Huang L, Appleton JA. Eosinophils in helminth infection: defenders and dupes. Trends Parasitol 2016;32:798–807. https://doi.org/10.1016/j.pt.2016.05.004.
[46] Lange AM, Yutanawiboonchai W, Scott P, Abraham D, Lange AM, Yutanawiboonchai W, et al. IL-4- and IL-5-dependent protective immunity to Onchocerca volvulus infective larvae in BALB/cBYJ mice. J Parasitol 1994:19104.
[47] Gazzinelli R, Eltoum I, Wynn TA, Sher A. Acute cerebral toxoplasmosis is induced by in vivo neutralization of TNF-alpha and correlates with the down-regulated expression of inducible nitric oxide synthase and other markers of macrophage activation. J Immunol 1993:151. https://doi.org/10.4049/jimmunol.173.6.3589.
[48] Nicholas WL, Stewart AC, Mitchell GF. Antibody responses to Toxocara canis using sera from parasite-infected mice and protection from toxocariasis by immunisation with es antigens. Aust J Exp Biol Med 1984;62(5):619–26. https://doi.org/10.1038/jcb.1984.59.
[49] Cuéllar C, Fenoy S, Águila C, Guillén JL. Isotype specific immune responses in murine experimental toxocariasis. Mem Inst Oswaldo Cruz 2001;96:549–53.
[50] Fan CK, Lin YH, Du WY, Su KE. Infectivity and pathogenicity of 14-month-cultured embryonated eggs of Toxocara canis in mice. Vet Parasitol 2003;113:145–55. https://doi.org/10.1016/S0304-4017(03)00046-3.
[51] Takamoto M, Wang Z-X, Watanabe N, Sugane K. The measurement of parasite antigen-specific IgE levels using anti-IgE monoclonal antibodies and biotinylated antigens. Parasitol Res 2001;87:919–23.
[52] Dvoroznakova E, Borošková Z, Dubinský P, Tomašovičová O, Machnicka B. Toxocara canis in mice: Immune responses after infection and immunization. Helminthologia 2000;37:199–204.
[53] Harris N, Gause WC. To B or not to B: B cells and the Th2-type immune response to helminths. Trends Immunol 2011;32:80–8. https://doi.org/10.1016/j.it.2010.11.005.
[54] Foster N, Elsheikha HM. The immune response to parasitic helminths of veterinary importance and its potential manipulation for future vaccine control strategies. Parasitol Res 2012;110:1587–99. https://doi.org/10.1007/s00436-012-2832-y.
[55] O'Brien LM, Fitzpatrick E, Baird AW, Campion DP. Eosinophil-nerve interactions and neuronal plasticity in rat gut associated lymphoid tissue (GALT) in response to enteric parasitism. J Neuroimmunol 2008;197:1–9. https://doi.org/10.1016/j.jneuroim.2008.04.002.
[56] Sher A, Fiorentino D, Caspar P, Pearce E, Mosmann T. Production of IL-10 by CD4+ T lymphocytes correlates with down-regulation of Th1 cytokine synthesis in helminth infection. J Immunol 1991;147:2713–6.
[57] Couper KN, Blount DG, Riley EM. IL-10: the master regulator of immunity to infection. J Immunol 2008;180:5771–7. https://doi.org/10.4049/jimmunol.180.9.5771.
[58] Schopf LR, Hoffmann KF, Cheever AW, Urban JF, Wynn TA. IL-10 Is critical for host resistance and survival during gastrointestinal helminth infection. J Immunol 2002;168:2383–92. https://doi.org/10.4049/jimmunol.168.5.2383.
[59] Maizels R. Infections and allergy - Helminths, hygiene and host immune regulation. Curr Opin Immunol 2005;17:656–61. https://doi.org/10.1016/j.coi.2005.09.001.
[60] Alcântara-Neves NM, Britto GDSG, Veiga RV, Figueiredo CA, Fiaccone RL, Conceição JS, et al. Effects of helminth co-infections on atopy, asthma and cytokine production in children living in a poor urban area in Latin America. BMC Res Notes 2014:1–11. https://doi.org/10.1186/1756-0500-7-817.
[61] Fallon PG, Mangan NE. Suppression of TH2-type allergic reactions by helminth infection. Nat Rev Immunol 2007;7:220–30. https://doi.org/10.1038/nri2039.
[62] Maizels R, Blaxter ML, Selkirk ME. Forms and functions of nematode surfaces. Exp Parasitol 1993;77:380–4. https://doi.org/10.1006/expr.1993.1096.
[63] Hewitson JP, Grainger JR, Maizels RM. Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity. Mol Biochem Parasitol 2009;167:1–11. https://doi.org/10.1016/j.molbiopara.2009.04.008.
[64] Piedrafita D, Spithill TW, Dalton JP, Brindley PJ, Sandeman MR, Wood PR, et al. Juvenile Fasciola hepatica are resistant to killing in vitro by free radicals compared with larvae of Schistosoma mansoni. Parasite Immunol 2000;22:287–95. https://doi.org/10.1046/j.1365-3024.2000.00300.x.

frontiers in Bioengineering and Biotechnology

BRIEF RESEARCH REPORT
published: 20 August 2019
doi: 10.3389/fbioe.2019.00200

Single-Input Regulatory Cascade for *in vivo* Removal of the Solubility Tag in Fusion Recombinant Proteins Produced by *Escherichia coli*

*Filipe S. R. Silva, Sara P. O. Santos, Roberto Meyer, Neuza M. Alcantara-Neves, Carina S. Pinheiro and Luis G. C. Pacheco**

Post-graduate Program in Biotechnology, Institute of Health Sciences, Federal University of Bahia, Salvador, Brazil

Solubility tags are commonly fused to target recombinant proteins to enhance their solubility and stability. In general, these protein tags must be removed to avoid misfolding of the partner protein and to allow for downstream applications. Nevertheless, *in vitro* tag removal increases process complexity and costs. Herein, we describe a synthetic biology-based strategy to permit *in vivo* removal of a solubility tag (EDA, KDPG aldolase), through co-expression of the fusion recombinant protein (EDA-EGFP) and the tag-cleaving protease (TEVp), in a controlled manner. Basically, the system uses three repressor proteins (LacI, cI434, and TetR) to regulate the expressions of EDA-EGFP and TEVp, in a regulatory cascade that culminates with the release of free soluble target protein (EGFP), following a single chemical induction by IPTG. The system worked consistently when all biological parts were cloned in a single plasmid, pSolubility(SOL)A (7.08 Kb, Amp^R), and transformed in *Escherichia coli* Rosetta (DE3) or BL21(DE3) strains. Total soluble recombinant protein yield (EDA-EGFP + free EGFP) was *ca.* 272.0 ± 60.1 μg/mL of culture, following IMAC purification; free EGFP composed great part (average = 46.5%; maximum = 67.3%) of the total purified protein fraction and was easily separated from remaining fusion EDA-EGFP (53 KDa) through filtration using a 50 KDa cut-off centrifugal filter.

Keywords: recombinant proteins, protein solubility, synthetic biology, *Escherichia coli*, green fluorescent protein

OPEN ACCESS

Edited by:
Pablo Ivan Nikel, The Novo Nordisk Foundation Center for Biosustainability (DTU Biosustain), Denmark

Reviewed by:
Michele Galluccio, University of Calabria, Italy
Xristo Zarate, Universidad Autónoma de Nuevo León, Mexico

****Correspondence:***
Luis G. C. Pacheco
luis.pacheco@ufba.br

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *05 February 2019*
Accepted: *06 August 2019*
Published: *20 August 2019*

Citation:
Silva FSR, Santos SPO, Meyer R, Alcantara-Neves NM, Pinheiro CS and Pacheco LGC (2019) Single-Input Regulatory Cascade for in vivo Removal of the Solubility Tag in Fusion Recombinant Proteins Produced by Escherichia coli. Front. Bioeng. Biotechnol. 7:200. doi: 10.3389/fbioe.2019.00200

INTRODUCTION

Fusion protein tags are normally used for successfully obtaining hard-to-express recombinant proteins in their soluble form in bacteria. A fusion tag can enhance a given recombinant protein quality by improving its translation, avoiding protein aggregation and even shielding it from degradation (Waugh, 2005; Kang et al., 2015; Bernier et al., 2018). Commonly used solubility enhancers include Maltose-binding protein (MBP, 42.5 KDa), Glutathione-S-transferase (GST, 26 KDa), Thioredoxin A (TrxA, 12 KDa), and N-utilization substance protein A (NusA, 55 KDa). Following expression of the fused recombinant protein, these protein tags need to be detached as they can significantly affect a given passenger protein's biological function. For this, specific protease cleavage sites are placed in between the fusion tag and the target protein, which can then be recovered in its natural form after *in vitro* incubation with the respective proteases, such as the Tobacco Etch Virus protease (TEVp), followed by chromatographic steps. However, these

post-processing steps increase production costs and process intricacy (Li, 2011). To circumvent these technical difficulties, some studies have tried to co-express the specific protease with the fusion protein to get the unfused target protein *in vivo* in a simpler manner (Kapust and Waugh, 2000; Shih et al., 2005; Wei et al., 2012; Feng et al., 2014; Luo et al., 2015). Generally, co-expression of TEVp with the fusion target protein is done by using different inducing agents (e.g., IPTG and aTc) (Kapust and Waugh, 2000), or by using the same operator site to control transcription of both genes (Wei et al., 2012). The protease can also be constitutively expressed through chromosomal integration, or transcriptionally fused to the cassette that codes for the fusion protein (Shih et al., 2005).

In this brief report, we propose a strategy based on a regulatory cascade to produce both the target fusion protein and the tag-cleaving protease TEVp through a single chemical induction, using different operator sites. Similarly, to the repressilator genetic circuit (Elowitz and Leibler, 2000), our system uses three repressor proteins (LacI, cI434, and TetR) to regulate the expression of the target fusion protein and the TEVp, in a regulatory cascade that culminates with *in vivo* release of EGFP from its solubility tag (**Figures 1A,B**).

MATERIALS AND METHODS

Genetic Circuit Design and Biological Parts Selection

The genetic elements used to compose the three genetic modules shown in **Figure 1A** were retrieved from the iGEM Registry of Standard Biological Parts (http://parts.igem.org/Main_Page) and from selected previous studies (**Supplementary Table S1**). The first module contains the T7 promoter, the *lacO* operator site and an RBS derived from the registry part # BBa_K567018. The sequence coding for a fusion target protein consisting of the solubility tag KDPG aldolase (EDA), a Gly-Ser-Gly-Ser flexible linker, a canonical TEVp cleavage recognition site (Glu-Asn-Leu-Tyr-Phe-Gln↓Gly) and EGFP, was then put under control of these genetic elements (**Figure 1A**). A 31 bp spacer sequence was placed upstream and an 8 bp spacer was situated downstream a medium strength RBS, which controls the translation of the cI434 repressor, that is transcriptionally coupled to the sequence encoding the fusion protein. The third module was designed to express the TetR repressor under control of the lambda promoter sequence, which is regulated by the cI434 repressor (**Figure 1A**). This way, TetR is expected to be produced when IPTG is absent in the growth medium (**Figure 1B**). Lastly, the TEVp is produced under the control of a TetR regulated promoter and translated using a weak RBS (**Figure 1A**). The repressor proteins have a C-terminal LVA degradation tail, which is expected to expedite degradation of these regulators in *Escherichia coli*, in order to prevent the circuit from collapsing due to the accumulation of regulators (Brophy and Voigt, 2014) (**Supplementary Figure S1A**).

Abbreviations: aTc, anhydrotetracycline; BCIP, 5-Bromo-4-chloro-3-indolyl phosphate; BME, 2-mercaptoethanol; cI434, phage 434 repressor protein; *cI434O*, operator site repressible by cI434; DTT, dithiothreitol; EDA, KHG/KDPG aldolase; EGFP, enhanced green fluorescent protein; iGEM, International Genetically Engineered Machine; IMAC, Immobilized metal affinity chromatography; IPTG, Isopropyl β-D-1-thiogalactopyranoside; KDPG, 2-Keto-3-deoxy-6-phosphogluconate; LVA, leucine - valin - alanine; NBT, nitroblue tetrazolium; OD, optical density; RBS, ribosome biding site; RFC23, BioBrick™ request for comments 23/Silver assembly; TEV, Tobacco Etch Virus.

Plasmids Design and Construction

The three modules were designed containing RFC23 BioBricks™ standard sites at extremities in order to facilitate assembling (**Figure 1A**) (Røkke et al., 2014). Synthetic constructions were purchased from GenScript (Scotch Plains, NJ, USA), initially cloned in pUC57 and then sub-cloned in BioBricks compatible plasmid backbones (**Supplementary Table S2**). Module 2 was isolated from pM2A vector by digesting it with *EcoRI* and *PstI* enzymes. Then, it was inserted into the predigested BioBricks compatible plasmid pSB1C3 to generate pM2C. To connect modules 1 and 2, Silver assembly (Phillips and Silver, 2006) was performed to join together EDA and EGFP coding sequences. For this, pM1A containing the EDA coding sequence was digested with *EcoRI* and *SpeI*, releasing the module 1 fragment (**Figure 1A**). On the other hand, pM2C was linearized with *EcoRI* and *XbaI*. The isolated module 1 and pM2C fragments were joined together using T4 DNA ligase (Promega), and the resulting plasmid was named pM12C. pM12C was then linearized with *SpeI* and *PstI* and pM3A was cut with *XbaI* and *PstI*. Following purification, these two fragments were ligated to the form pSOLC, which contains the three modules. pSOLC was digested with *EcoRI* and *PstI* and then inserted back in pUC57, resulting in the plasmid pSOLA. Finally, pM3A was also digested with *EcoRI* and *PstI* and cloned into pSB1K3 to give pM3K (http://partsregistry.org/Part:pSB1C3). Plasmids constructions are summarized in **Figure 1C**. Details are given on **Supplementary Methods**.

Recombinant Protein Expression, Purification, and Analysis

Chemically transformed *E. coli* strains (BL21, Rosetta™, and CodonPlus-RIL), were routinely maintained at 37°C, with aeration, in Luria-Bertani (LB) broth or LB-agar plates, containing the appropriate antibiotics according to the plasmid-conferred resistances (pSOLA/Amp^R; pSOLC/Cm^R; pM12C + pM3K/Cm^R + Kan^R). Recombinant protein production was induced by the addition of 0.5 mM IPTG to growth media, when cells reached optical densities (at 600 nm) of 0.6, 1.5, or 3.0; bacterial cultures were further incubated at 25°C for up to 24 h. Fluorescence emission by recombinant expression of EGFP in cultures was monitored by Fluoroskan Ascent™ Microplate Fluorimeter (Ex. = 485 nm; Em. = 535 nm). Aliquots were collected at different time points, bacterial pellets were lysed by sonication in FastBreak™ Cell Lysis Reagent (Promega, Madison, WI, USA), and total protein extracts were analyzed by 12% SDS-PAGE (250 mM of DTT or BME) and Western blotting using eGFP Tag Monoclonal Antibody (Invitrogen, F56-6A1.2.3, 1:4000). IMAC protein purification of 6xHis-tagged recombinant proteins was performed using MagneHis™ (Promega, Madison, WI, USA). Additionally, the recovered purified protein fraction was filtered through

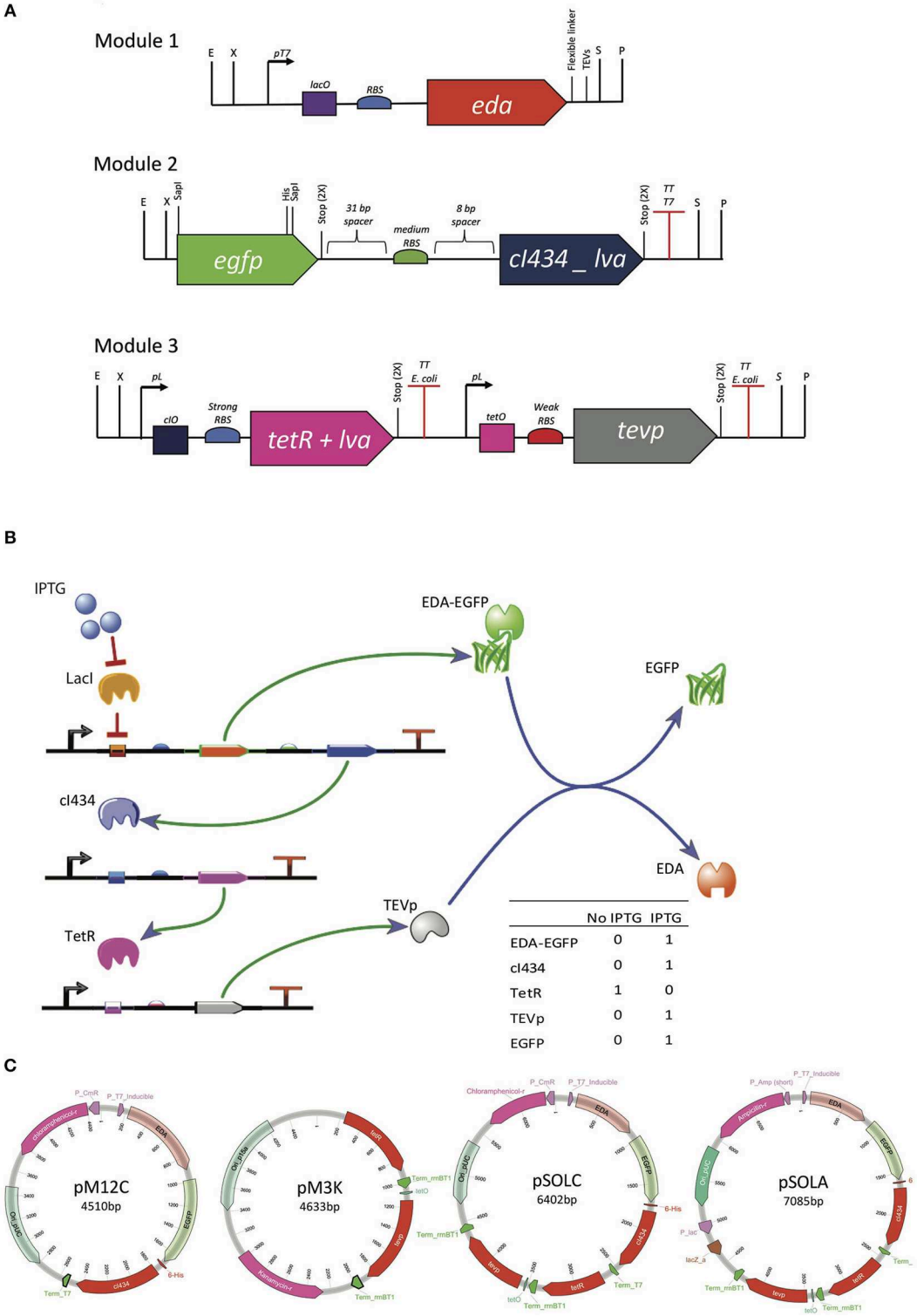


FIGURE 1 | Genetic organization of the system for controlled intracellular processing of recombinant proteins. **(A)** Genetic modules built with biological parts described in **Supplementary Table S1**, synthetized with RFC23 Biobrick standard, to allow for easy assembly. **(B)** Genetic circuit graphic simulation, built with TinkerCell (Chandran et al., 2009). **(C)** Plasmids assembled from the tree different modules. Modules were distributed in two different plasmids (pM12C + pM3K) or joined in one plasmid (pSOLA or pSOLC). pM12C contains both modules 1 and 2 joined together and has pSB1C3 (high copy, Cm^R) backbone. pM3K has the module 3 in a pSB1K3 (low copy, Km^R) backbone. pSOLC includes all three modules inserted in pSB1C3 (high copy, Cm^R) and pSOLA holds all three modules introduced in pUC57 backbone (high copy, Amp^R).

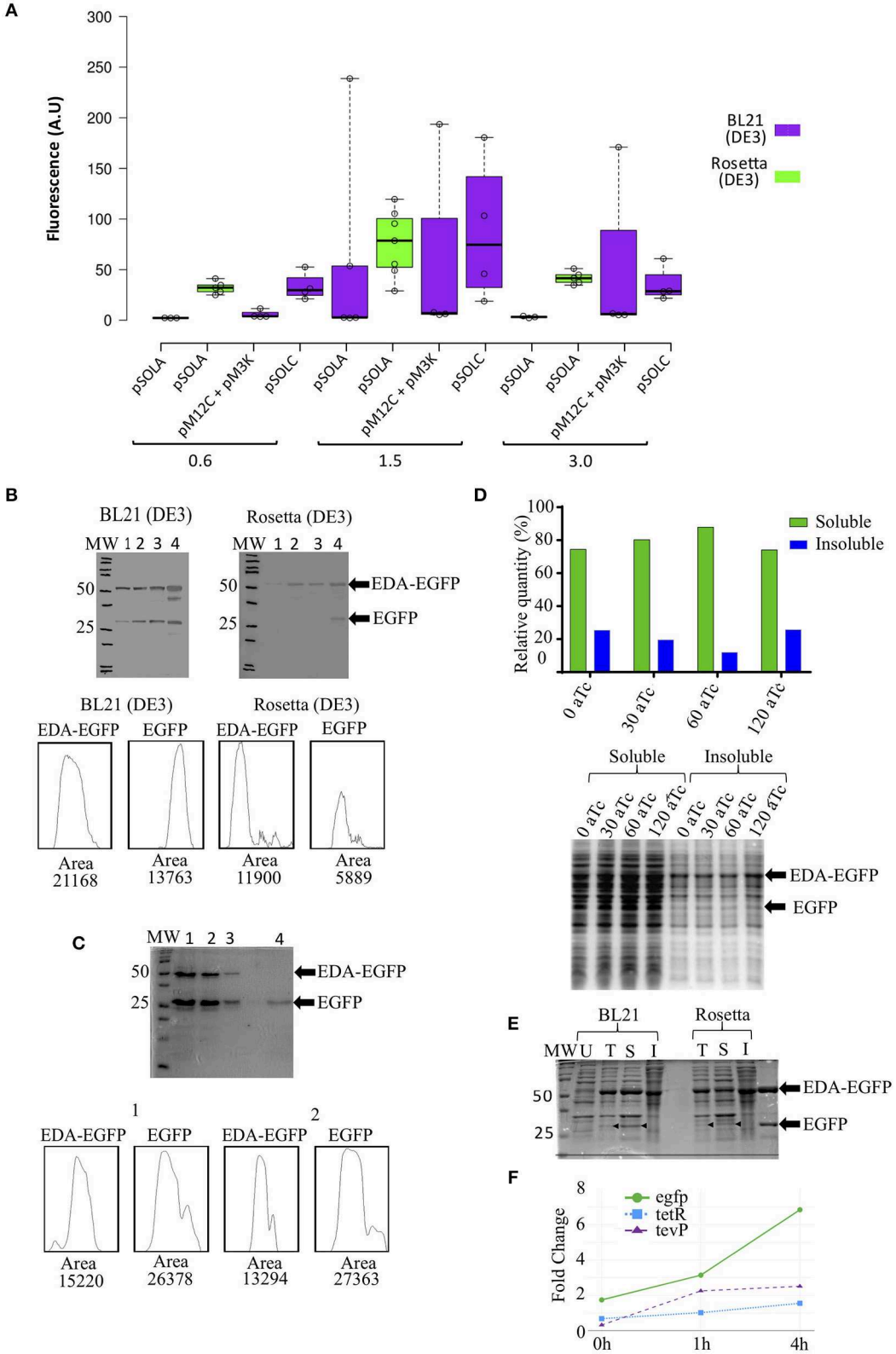


FIGURE 2 | EDA-EGFP fusion protein production and levels of released EGFP, using different genetic organizations and *E. coli* strains. **(A)** Fluorescence measurements after 24 h of IPTG induction at different optical densities (OD_{600nm}); **(B)** Western blot detection of EDA-EGFP and EGFP proteins by anti-eGFP Tag

(Continued)

FIGURE 2 | Monoclonal Antibody at: (1) 0 h, (2) 4 h, (3) 6 h, and (4) 24 h post induction. Densitometric analyses of Western blot detections (lane 4) is presented; 30 μg of each sample was loaded per each well; **(C)** 15% SDS-PAGE from protein purification of untagged EGFP. Below are densitometric analysis of 24 h protein profile for both BL21 and Rosetta by ImageJ software. (1 and 2) 6xHis-tagged proteins obtained following IMAC purification (MagneHis^{TM} Protein Purification System, soluble protein protocol); (3) Retained protein concentrate in the filter Amicon^{TM} Ultra (>50 KDa); (4) collected EGFP fraction (flow-through) (<50 KDa). Densitometric analysis by ImageJ of lanes 1 and 2 are also shown. **(D)** Relative quantification of EGFP found in the soluble and insoluble fractions (from cultures containing 0, 30, 60, and 120 ng/mL of anhydrotetracycline) in SDS-PAGE after solubility test. **(E)** 12% SDS-PAGE of (U) uniduced cells; (T) total lysated cells and (S) soluble and (I) insoluble fractions from solubility test. Thirty microgram of total proteins were loaded in each lane. **(F)** Relative gene expression analysis of the three transcriptional units in Rosetta(DE3) at 0, 1, and 4 h after induction with IPTG. Primers used are listed in **Supplementary Table S4**.

Amicon^{TM} (Lexington, MA, USA) Ultra 2 mL Centrifugal Filter (50 kDA cut-off) (see **Supplementary Methods** for details).

RESULTS AND DISCUSSION

Figure 1 shows the genetic organization of the system for controlled intracellular processing of a recombinant fusion protein, in order to release the solubility tag *in vivo* with a single chemical induction. The expected functioning of the system is the following: upon IPTG induction, the target fusion protein (EDA-EGFP) is produced along with the cI434 repressor; cI434 in turn binds to its cognate operator site and stops TetR production; TEV protease, which is repressed by binding of TetR to *tetO* operator site, then starts to be produced (**Figure 1B**; **Supplementary Figure S1A**). The genetic modules were all cloned in a single plasmid (pSOLA or pSOLC, for Amp^R and Cm^R, respectively) or in two different plasmids (pM12C + pM3K, Cm^R and Kan^R), with differing copy numbers, in order to tune the production of the various components at their required levels (**Figures 1A,C**; **Supplementary Table S2**).

BL21(DE3) *E. coli* cells carrying pSOLA (which has the three genetic modules in a single plasmid) rendered the highest EGFP fluorescence signal among all tested conditions, when IPTG induction was added at an OD_{600nm} = 1.5 (mean fluorescence units FU = 60.01 ± 102.30 A.U.; maximum FU = 238.70 A.U), though fluorescence levels were highly variable in this strain (**Figure 2A**); significant increase in recombinant protein expression was reached at 24 h post-induction in this strain (**Supplementary Figure S2**). *E. coli* Rosetta (DE3) in turn showed a more reproducible EGFP fluorescence signal generation throughout all replicates, despite reaching apparently lower induction levels (mean FU = 76.09 ± 32.89 A.U.; maximum FU = 119.50 A.U.) (**Figure 2A**); noteworthy, these fluorescence values were not significantly different from other induction conditions at OD_{600nm} = 1.5, indicating that the system works similarly in both strains (**Figure 2A**; **Supplementary Table S3** and **Figure S2**). Rosetta (DE3) can be used to overcome low yield and poor solubility of recombinant TEVp produced in *E. coli* (Wei et al., 2012; Cesaratto et al., 2016). We hypothesize that this might be a contributing factor for obtaining more predictable results using this strain in this study.

While EGFP accumulation was also observed in the two-plasmid based system (pM12C + pM3K) using *E. coli* BL21 (DE3), no TEVp activity was detected *in vivo* (not shown); conversely, the fused EDA-EGFP (53 KDa) protein was completely cleaved *in vitro* with purified recombinant TEVp, releasing the EDA tag (23 KDa) and the his-tagged recombinant EGFP (30 KDa) (**Supplementary Figure S2D**). The single-plasmid based system (pSOLA) in turn, showed significant *in vivo* cleavage of the EDA-EGFP fusion protein for both BL21 (DE3) and Rosetta (DE3) (**Figure 2B**). Significant leaking is observed when EDA-EGFP is expressed in BL21(DE3); besides, released EGFP can be found at early induction times, but accumulates at higher concentrations at 24 h post-induction. Conversely, expression in Rosetta (DE3) was closer to what would be expected from the genetic system functioning (**Figure 2B**); this is also confirmed by gene expression analysis of the three transcriptional units that compose the system, which shows higher expression of *egfp* when compared to *tevP*, in all time points (**Figure 2F**; **Supplementary Figure S1B**). Novel combinations of biological parts can be tested in future constructions to evaluate their effects on fine-tuning of the genetic system. This will be important to address a limitation of our approach, that was the persistence of significant part of the recombinant protein still in its fusion form *in vivo* (**Figures 2B,E**), whereas previous studies of controlled intracellular processing in *E. coli* have achieved almost complete processing of solubility tags (Kapust and Waugh, 2000; Nallamsetty et al., 2004; Raran-kurussi and Waugh, 2016).

Soluble 6xHis-tagged proteins were purified using a bead-based protocol, yielding *ca.* 272.0 ± 60.1 μg/mL of purified recombinant EDA-EGFP and free EGFP per mL of culture (**Figure 2C**; **Supplementary Figure S2**). Densitometric analysis showed variable proportions of EDA-EGFP/EGFP, ranging from 0.3- to 2-fold concentration of untagged protein compared to EDA-tagged protein (**Figure 2C**; **Supplementary Figure S2**). Purified protein was then submitted to diafiltration using a 50 KDa cut-off centrifugal filter, in order to separate fusion EDA-EGFP from detached EGFP (**Figure 2C**; **Supplementary Figure S2**). In future configurations, EDA can be substituted by another solubility partner such as MBP, then permitting removal by affinity chromatography (Kosobokova et al., 2016).

Increasing concentrations of anhydrotetracycline (aTc) were added to the culture media after 4 h of IPTG induction to check whether it would enhance *in vivo* protein cleavage, as described by Kapust and Waugh (2000). The proportion of soluble EGFP recovered was around 80.0% of total recombinant EGFP protein produced, either with no addition of aTc or with aTc concentrations ranging from 30 to 120 ng/mL (**Figure 2D**). These results indicate that only IPTG induction is sufficient to simultaneously express the fusion protein and TEVp, resulting in untagged EGFP in absence of anhydrotetracycline. **Figure 2E** shows that released EGFP is found mostly in the soluble fraction.

The genetic regulatory cascade described here is composed by genetic elements that interact among themselves resulting in

the simultaneous production of a fusion recombinant protein and of the site-specific protease that separates the solubility tag from the target protein, all with a single induction. The main characteristics of this genetic system are: (i) it requires only a single inducing agent (IPTG); (ii) it is tuned to produce a higher amount of the fusion recombinant protein than the tag-cleaving protease; (iii) it can potentially be adapted to any cell lineage that produces T7 RNA polymerase. This genetic circuit is able to perform the task of co-producing both EDA-EGFP fusion protein with tag-cleaving TEVp, then resulting in an average of 46.5% (maximum 67.3%) of soluble EGFP release *in vivo* (**Figure 2C**).

AUTHOR CONTRIBUTIONS

FS and SS conducted all the experiments. RM, NA-N, CP, and LP conceived experiments, discussed results, and contributed to manuscript writing. All authors read and corrected the final manuscript.

FUNDING

This research was supported by the following research grants: FAPESB/CNPq—PRONEM PNE 007/2014; and MCTI/CNPq/FNDCT—RENORBIO 79/2013. FS was recipient of a scholarship from FAPESB. NA-N and LP were recipients of research fellowships from CNPq.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2019.00200/full#supplementary-material

REFERENCES

Bernier, S. C., Cantin, L., and Salesse, C. (2018). Systematic analysis of the expression, solubility and purification of a passenger protein in fusion with different tags. *Protein Expr. Purif.* 152, 92–106. doi: 10.1016/j.pep.2018.07.007

Brophy, J. A., and Voigt, C. A. (2014). Principles of genetic circuit design. *Nat. Methods* 11, 508–520. doi: 10.1038/nmeth.2926

Cesaratto, F., Burrone, O. R., and Petris, G. (2016). Tobacco Etch virus protease: a shortcut across biotechnologies. *J. Biotechnol.* 231, 239–249. doi: 10.1016/j.jbiotec.2016.06.012

Chandran, D., Bergmann, F. T., and Sauro, H. M. (2009). TinkerCell: modular CAD tool for synthetic biology. *J. Biol. Eng.* 3, 1–17. doi: 10.1186/1754-1611-3-19

Elowitz, M. B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338. doi: 10.1038/35002125

Feng, Y., Xu, Q., Yang, T., Sun, E., Li, J., Shi, D., et al. (2014). A novel self-cleavage system for production of soluble recombinant protein in *Escherichia coli*. *Protein Expr. Purif.* 99, 64–69. doi: 10.1016/j.pep.2014.04.001

Kang, Y. S., Song, J. A., Han, K. Y., and Lee, J. (2015). *Escherichia coli* EDA is a novel fusion expression partner to improve solubility of aggregation-prone heterologous proteins. *J. Biotechnol.* 194, 39–47. doi: 10.1016/j.jbiotec.2014.11.025

Kapust, R. B., and Waugh, D. S. (2000). Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr. Purif.* 19, 312–318. doi: 10.1006/prep.2000.1251

Kosobokova, E. N., Skrypnik, K. A., and Kosorukov, V. S. (2016). Overview of fusion tags for recombinant proteins. *Biochemistry* 81, 187–200. doi: 10.1134/S0006297916030019

Li, Y. (2011). Self-cleaving fusion tags for recombinant protein production. *Biotechnol. Lett.* 33, 869–881. doi: 10.1007/s10529-011-0533-8

Luo, X., Li, L., Chai, M., Zhang, Q., and Shang, G. (2015). *Escherichia coli* BL21(DE3) chromosome-based controlled intracellular processing system for fusion protein separation. *J. Microbiol. Methods* 114, 35–37. doi: 10.1016/j.mimet.2015.04.013

Nallamsetty, S., Kapust, R. B., Tözsér, J., Cherry, S., Tropea, J. E., Copeland, T. D., et al. (2004). Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease *in vivo* and *in vitro*. *Protein Expr. Purif.* 38, 108–115. doi: 10.1016/j.pep.2004.08.016

Phillips, I., and Silver, P. (2006). *A New Biobrick Assembly Strategy Designed for Facile Protein Engineering*. Cambridge, MA: Massachusetts Institute of Technology. Available online at: http://hdl.handle.net/1721.1/32535

Raran-kurussi, S., and Waugh, D. S. (2016). A dual protease approach for expression and af fi nity purification of recombinant proteins. *Anal. Biochem.* 504, 30–37. doi: 10.1016/j.ab.2016.04.006

Røkke, G., Korvald, E., Pahr, J., Oyås, O., and Lale, R. (2014). BioBrick assembly standards and techniques and associated software tools. *Methods Mol. Biol.* 1116, 1–24. doi: 10.1007/978-1-62703-764-8_1

Shih, Y. P., Wu, H. C., Hu, S. M., Wang, T. F., and Wang, A. H. (2005). Self-cleavage of fusion protein *in vivo* using TEV protease to yield native protein. *Protein Sci.* 14, 936–941. doi: 10.1110/ps.041129605

Waugh, D. S. (2005). Making the most of affinity tags. *Trends Biotechnol.* 23, 316–320. doi: 10.1016/j.tibtech.2005.03.012

Wei, L., Cai, X., Qi, Z., Rong, L., Cheng, B., and Fan, J. (2012). *In vivo* and *in vitro* characterization of TEV protease mutants. *Protein Expr. Purif.* 83, 157–163. doi: 10.1016/j.pep.2012.03.011

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.